UNIVERSITAT DE BARCELONA

Epigenetic determinants of differentiation in monocyte subsets: implications in autoimmune disease

Anna Guiomar Ferreté Bonastre

EPIGENETIC DETERMINANTS OF DIFFERENTIATION IN MONOCYTE SUBSETS: IMPLICATIONS IN AUTOIMMUNE DISEASE

Memòria presentada per Anna Guiomar Ferreté Bonastre per optar al títol de doctora per la Universitat de Barcelona

2024

UNIVERSITAT DE BARCELONA - FACULTAT DE FARMÀCIA
PROGRAMA DE DOCTORAT EN BIOMEDICINA

Tesis realitzada al grup d'Epigenètica i Malalties Immunitàries de l'Institut de Recerca contra la Leucèmia Josep Carreras sota la direcció del Dr. Esteban Ballestar Tarín

Dr. Esteban Ballestar	Dra. Carme Caelles	Anna G. Ferreté Bonastre
Director	Tutora	Doctorand

I. INDEX

II. ABSTRACT

Monocytes are highly plastic and heterogeneous immune cells that play fundamental roles in the innate immune system. Single-cell omic studies have revealed a broad range of unique phenotypes and behaviors in the cells comprising this myeloid population, which has historically been studied as a unique cell type. To study them through bulk techniques, monocyte cells are grouped into smaller subpopulations to simplify their heterogeneity. The most common classification generates three non-overlapping subsets based on the expression of the surface molecules CD14 and CD16. The subsets receive the names of classical (cMO), intermediate (iMO) and non-classical monocytes (ncMO) and they originate via a sequential differentiation process in the reported order. The mechanisms leading their differentiation progress in humans are unknown. Each subset has been attributed distinct immune functions. Also, differences in the proportions of monocyte subsets are reported in multiple physiological and pathological conditions. However, specific contributions of the monocyte subsets to disease development have rarely been explored. In this thesis, the molecular mechanisms regulating human monocyte subset differentiation were explored. Additionally, dysregulation of monocyte subsets in relation to the archetypical autoimmune disease Systemic Lupus Erythematosus (SLE) was also inspected. Both studies applied a combination of bulk epigenomic and transcriptomic analysis of monocyte subsets paired with the integration of exceptional single-cell RNA-seq datasets.

Firstly, the differentiation process of monocyte subsets was studied in steady state. DNA methylome and transcriptome analysis of the three monocyte subsets, was consistent with their subsequent linear differentiation process. Gene ontology analyses of the differentially methylated positions and the differentially expressed genes among monocyte subsets were also consistent with immunological functions previously ascribed to each subset in the literature. As anticipated, we described a significant correlation between methylation levels at certain positions and expression of the associated genes, particularly for positions located at promoters and intronic regions. Most surprisingly, we identified two transcription factors throughout

monocyte subset differentiation: HIF-1α and T-bet. HIF-1α and its target genes were significantly less methylated and more expressed in cMOs. T-bet, which has never before been reported to be expressed in unstimulated monocytes, exhibited an opposing behavior by being predominant in ncMOs. Integration with single cell transcriptomics data from a patient with a homozygous mutation for the gene encoding T-bet supported a general role of this transcription factor in the development of the myeloid compartment and the specific contribution to ncMO differentiation. In this patient, functions previously attributed to ncMOs, such as TNF secretion, appeared affected by the mutation. Similarly, HIF-1α signature was aberrantly regulated in the mutant, confirming our previously proposed interaction between the two factors.

Secondly, we investigated the epigenomics and transcriptomics alterations of the three monocyte subsets in the autoimmune disease SLE. We collected a prospective cohort of over 20 SLE patients at the onset of a flare episode as well as samples from 10 healthy controls. DNA methylation profiles of the monocyte subsets from these samples revealed the predisposition of each subset to respond differentially to the inflammatory environment present in the disease. cMOs presented a high association of demethylation patterns with the family of transcription factors Fos/Jun. These results suggest propensity to macrophagic differentiation from these cells. ncMOs presented fewer specific differences but they were directed towards phenotypes best described in T lymphocytes, indicating a potential strengthened interaction between this subset and T cells. iMOs phenotypes were recurrently observed between the other two subsets. Transcriptomic profiling revealed cMOs as the subset with higher capability to respond to the inflammatory milieu of SLE, probably due to their less differentiated initial state. We also described an association of the three subsets' datasets with disease activity, particularly regarding the methylation and expression of components of the STAT pathway. Additionally, we reported three CpGs that could be used as prognosis marker predicting the outcome of the flare episode. Finally, integration with a large single-cell RNA-seq dataset revealed the high heterogeneity of SLE samples and how the composition of our bulk samples was enriched by pathological cell subdivisions.

III. ABBREVIATIONS

APC	antigen presenting cells
cMOs	classical monocytes
DAMPs	damage-associated molecular patterns
DEG	differentially expressed gene
DMP	differentially methylated position
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dsDNA	double-stranded DNA
FC	fold change
FDR	false discovery rate
GAS	γ-interferon activated site
GMP	granulocyte and macrophage progenitor
GO	gene ontology
GSEA	gene set enrichment analysis
HSC	hematopoietic stem cell
IFN	interferon
IFNAR	interferon-α receptor
IFNGR	interferon-γ receptor
iMOs	intermediate monocytes
IRF	interferon-regulatory factor
ISRE	interferon-stimulated response elements
JAK1	Janus kinase 1
LPS	lipopolysaccharide
MDP	monocyte and dendritic cell progenitor
MHC	major histocompatibility complex, (-I/-II: class I / class II)
ncMOs	non-classical monocytes
NK	natural killer
PAMPs	pathogen-associated molecular pattern molecules

PBMCs	peripheral blood mononuclear cells
PBS	phosphate buffered saline
RNA	ribonucleic acid
SAM	S-adenosylmethionine
SLE	systemic lupus erythematosus
SLEDAI	systemic lupus erythematosus disease activity index
SNP	single nucleotide polymorphism
STAT	signal transducer and activator of transcription
TET	ten-eleven translocation (referred to the TET family of 5-methylcytosine dioxygenases)
TF	transcription factor
TLR	toll-like receptor
TSS	transcription start site
TYK	tyrosine kinase
VST	variance stabilizing transformative

1 INTRODUCTION

1.1 The immune system

The development of immune systems has been essential for the survival and adaptability of organisms throughout evolution. The capacity to discern between beneficial and pathogenic interactions has allowed the creation of symbiotic engagements while repelling damaging insults. In general terms, the human immune system is divided into two categories: the innate immune system and the adaptive immune system.

The innate immune system is generally conceived as the first line of defense against pathogens and is shared by all multicellular entities. It offers a rapid response against pathogens and foreign substances recognizing conserved antigens or PAMPs (pathogen-associated molecular pattern molecules). In humans, it consists of the permanent physical barriers provided by the skin and mucous membranes, as well as the protection conferred by the cells and proteins constituting the innate immune system [1]. Overall, the innate immune system responds uniformly to all pathogens in a nonspecific manner. These responses consist in recognizing and destroying the pathogen through an inflammatory response and phagocytosis, i.e. the engulfment of the invader by phagocytic cells. The primary cells of the innate immune system in humans include neutrophils, basophils, eosinophils, mast cells, natural killer cells and, of particular focus in this thesis, monocytes, which differentiate into macrophages and dendritic cells.

The adaptive immune system emerged in evolution less than 500 million years ago [2–4], and is exclusive to vertebrate organisms. Its primary distinction from the innate immune system is its capacity to mount specific responses to different types of threats and retain memory of previous encounters with similar pathogens. In this way, adaptive immune cells are able to respond more quickly and efficiently upon posterior contacts. The adaptive immune system is intricately linked with and regulated by the innate immune system. It is mainly composed by T lymphocytes and B lymphocytes which further differentiate into numerous specialized subtypes of immune cells.

The cells forming the immune system and their self-renewing progenitors collectively form the hematopoietic system. Historically, hematopoietic differentiation has been understood as a stepwise process, with defined initial progenitor cell phenotypes that acquire some of the posterior progenitor phenotypes upon cell commitment. Recent advances in single-cell technologies have revealed this process to be gradual and continuous. What was previously categorized as a phenotype of a progenitor is now understood to encompass various individual cell types that partially share some features but not others. Nevertheless, it is still useful and common practice to classify the groups of progenitor cells according to previously established definitions [5,6]. In brief, at the top of the hierarchical pyramid of progenitors and immune cells are hematopoietic stem cells (HSC) with multi-potency and self-renewal capacity. HSC give rise to common myeloid and lymphoid progenitors with restricted multi-potent capabilities. Common myeloid progenitors give rise to monocytes and other cells of the innate immune system, which will be discussed shortly.

Common lymphoid progenitor cells yield T cells, B cells and natural killer (NK) cells, with the former two forming the foundation of the adaptive immune system. T and B cells have the potential to recognize and react to virtually any antigen, both pathogenic and benign. In consequence, they must undergo stepwise regulatory processes to eliminate the lymphocytes recognizing non-harmful antigens and retain the clones recognizing harmful, pathogenic antigens. These processes constitute a crucial characteristic for organism survival known as immunological tolerance. Posteriorly, lymphocytes can specialize into effector cells. B cells mature into antibody producing cells, while T cells can differentiate into cytotoxic cells or helper cells. In turn, helper T cells can differentiate into pro-inflammatory (Th1 and Th17), anti-inflammatory (Th2) or regulatory (Treg) cells, among others, depending on the cytokines they specialize in secreting. Finally, despite their lymphoid origin, NK cells are considered part of the innate immune system because they lack antigen-specific receptors. They recognize and eliminate infected, stressed and cancerous cells.

1.2 Differentiation of the myeloid lineage

Most cells of the innate immune system are derived from common myeloid progenitors through a series of hierarchical differentiations. In each step of the differentiation, the progenitors lose pluripotency capacity and gain resemblance to their final phenotype. In this process, megakaryocyte, the cells giving rise to platelets, and erythrocytes are among the earliest to differentiate from initial myeloid progenitors. Granulocyte, including basophils, eosinophils and neutrophils, all originate from the same pluripotent progenitor called granulocyte and macrophage progenitor (GMP). In parallel, GMP can also yield monocyte and dendritic cell progenitors (MDP) which then can differentiate into either monocytes or several types of dendritic cells [6,7] (Figure 1). All these differentiation steps take place in the bone marrow and the differentiated cells must egress to the bloodstream to perform their functions.

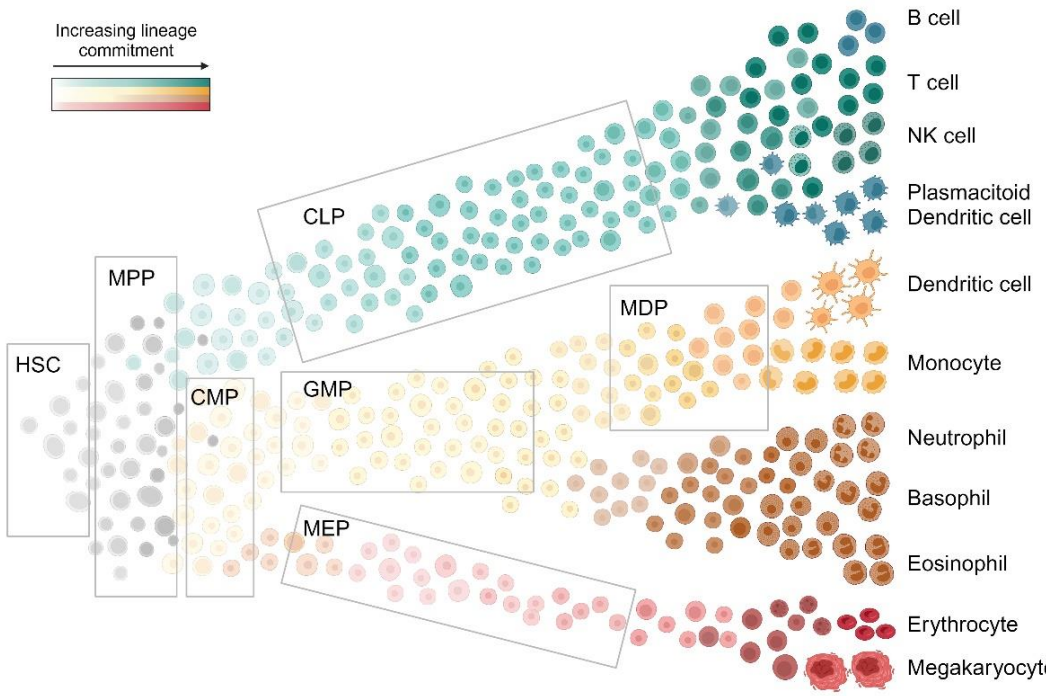


Figure 1. Hematopoietic system differentiation process. Scheme of the hierarchical gradient differentiation process of the cells constituting the hematopoietic system in the bone

marrow. Increase in color gradient represents the continuous rise in lineage commitment as cells divide and lose differentiation potency. HSC, hematopoietic stem cells; MPP, multipotent progenitor; CLP, common lymphocyte progenitor; CMP, common myeloid progenitor; GMP, granulocyte and macrophage progenitor; MEP, megakaryocyte and erythroid progenitor; MDP, macrophage and dendritic cell progenitor. Adapted from [5,6,8].

Macrophages are also a major component of the myeloid lineage of immune cells. Although some types of macrophages can differentiate from monocytes, the majority of tissue-resident macrophages of an adult human do not originate from monocytes egressed from peripheral blood. Instead, they have an embryonic and fetal root and populate the tissues before birth. They are maintained throughout adult life by their own self-renewing capacity [9]. The homeostatic contribution of monocytes to tissue resident macrophages is restricted to only a few tissues (gut, dermis and heart) or to inflammatory conditions.

Certainly, monocytes retain some, although limited, differentiation potential. Under regular conditions, they can differentiate to monocyte-derived macrophages and monocyte-derived dendritic cells. Both cell types can have either proinflammatory or tolerogenic properties which are acquired under the proper differentiation stimuli. For example, monocytes treated with granulocyte-macrophage colony stimulating factor (GM-CSF) will differentiate to macrophages with a proinflammatory phenotype, receiving the name of M1-like macrophages. However, in the presence of the cytokine macrophage colony-stimulating factor (M-CSF), they will differentiate to an M2-like phenotype, which is anti-inflammatory [10,11]. Moreover, under specific conditions, monocytes can also yield highly specialized cells such as Langerhans cells, microglia, and osteoclasts, critical in bone resorption.

Dendritic cells and macrophages constitute two of the most important professional antigen-presenting cells (APCs). APCs specialize in internalizing and digesting extracellular pathogens that have entered the body. Posteriorly, they expose fragments of these pathogens in their surface via the surface molecules named major histocompatibility complex of class II (MHC-II). This presentation is

recognized by T cells and can directly influence their activation and differentiation. APCs constitute a major link between innate and adaptive immune systems [12].

1.3 Monocytes

Monocytes are a short-lived and plastic cell type of the innate immune system. They compose around 10% of the peripheral blood mononuclear cells (PBMCs). Far from being mere APC precursors, monocytes make their own contributions to the immune response. Contrary to the differentiated cells they provide, which reside in tissues, monocytes are mainly found in the bloodstream and it is in circulation that they perform the majority of their functions.

Although historically regarded as a single cell type, monocytes are now recognized as a heterogeneous group of cells, as observed by the recent advancements in single-cell techniques such as flow cytometry and, especially, single-cell omics. The most common strategy to subdivide monocytes into smaller cell subtypes is by the expression of two cell surface markers: CD14 and CD16.

- CD14 was initially described as a marker for monocytes and macrophages [13] and later identified to be a pattern recognition receptor of PAMPs of bacterial origin, including lipopolysaccharide (LPS) [14]. In the last decades, new roles for CD14 have emerged relating it to the regulation of metabolism and insulin resistance [15]. It has also been linked to the development of autoimmune diseases through the recognition of damage-associated molecular patterns (DAMPs) such as those accumulated upon organ injury [16].
- CD16, encoded by the gene *FCGR3A*, is a low-affinity immunoglobulin receptor. It is best known for its expression in NK cells where it mediates an antibody-dependent cellular cytotoxicity. It is also expressed in some monocytes and macrophages, where it performs a similar role [17], facilitating the recognition and elimination of both cancer cells and virus-infected cells.

By the expression of these surface markers, monocytes are divided into three subsets: CD14+ CD16- classical monocytes, CD14+ CD16+ intermediate monocytes and CD14dim CD16+ non-classical monocytes (Figure 2).

1.3.1 Classical monocytes

Classical monocytes, henceforth termed cMOs, constitute the main subset of monocytes, comprising up to 85% of the total monocyte population. This subpopulation is considered to be the precursor, less differentiated, of the other two subsets because of its unique presence in the bone marrow where it originates from myeloid progenitors. According to an *in vivo* study by Patel et al. in human volunteers [18], cMOs only live for up to one day in the bloodstream. Posteriorly, they either egress to tissues to differentiate into macrophages or dendritic cells, differentiate within the bloodstream into intermediate monocytes, or die. Owing to their less differentiated features, they are thought to be more plastic than the rest of the subsets. Some studies report them able to equally differentiate to both dendritic cells and macrophages, contrary to the other subpopulations of monocytes which tend to favor macrophagic differentiation [19].

cMOs are considered to be the most pro-inflammatory out of the three subsets because of their highest expression of inflammatory molecules upon activation of most PAMP and DAMP receptors, known as Toll-like receptors (TLRs) [19]. Also due to their higher expression of pro-inflammatory S-100 proteins, which confer them the ability to support inflammation [20]. It is generally agreed that they are the responding subset for molecules of bacterial origin through the production of several pro-inflammatory cytokines such as CCL2, CCL3, IL-6 and IL-8 [21]. Furthermore, upon differentiation into macrophages, they are the subset that displayed the highest phagocytic capacity in some studies [19]. Conflicting with these premises, cMOs are also reported as a major producer of the anti-inflammatory cytokine IL-10 in response to LPS [20–22], which paints this subset with regulatory roles. Other functions attributed to cMOs include wound healing, tissue repair and migration to tissues [6].

1.3.2 Intermediate monocytes

Intermediate monocytes, hereafter referenced as iMOs, represent the least abundant subset in physiological conditions, corresponding to only 5-7% of the total fraction of monocytes. They originate within the bloodstream differentiating from a small fraction of cMOs [6]. They might persist in circulation for up to four days before extravasation, death or differentiation towards non-classical monocytes. Historically, iMOs are the least studied subset because of being considered as simply a transitional state in the differentiation process from cMOs to non-classical monocytes. This assumption is supported by the fact that they present a continuum intermediate expression of several surface markers as well as of the 90% of highly expressed genes [20]. However, the existence of some genes and surface markers specific to this subset suggests that they have individual and differentiated properties separating them from being a plain translational population [20,23].

iMOs have the highest levels of MHC-II molecules in comparison to the other monocyte subsets [20,23]. This, together with functional reports [24], proposes that they significantly participate in antigen processing and presentation processes as well as in interaction and activation of T cells [20]. Moreover, their genetic profile indicates a potential role in angiogenesis, although this can also be implied for the other subsets [25]. iMOs might as well have an anti-inflammatory role by means of production of IL-10. In fact, some reports depict iMOs as the main producers of this anti-inflammatory cytokine in response to LPS [26].

1.3.3 Non-classical monocytes

Non-classical monocytes, ncMOs hereafter, comprise around 10% of the total of monocytes under physiological conditions. They originate within the bloodstream from a fraction of iMOs and they are the longest lived of the monocyte subpopulations with an average lifespan of seven days [27]. Although they are characterized by a further differentiated phenotype than the other subsets, they retain certain capacity to terminally differentiate into macrophages and dendritic cells. Admittedly, without the presence of differentiation factors, ncMOs are the most susceptible monocyte

subset to differentiate to macrophages [19]. Additionally, some studies have shown that macrophages differentiated from ncMOs exhibit higher phagocytic capacity than those differentiated from cMOs, which contradicts the previously mentioned results. [28].

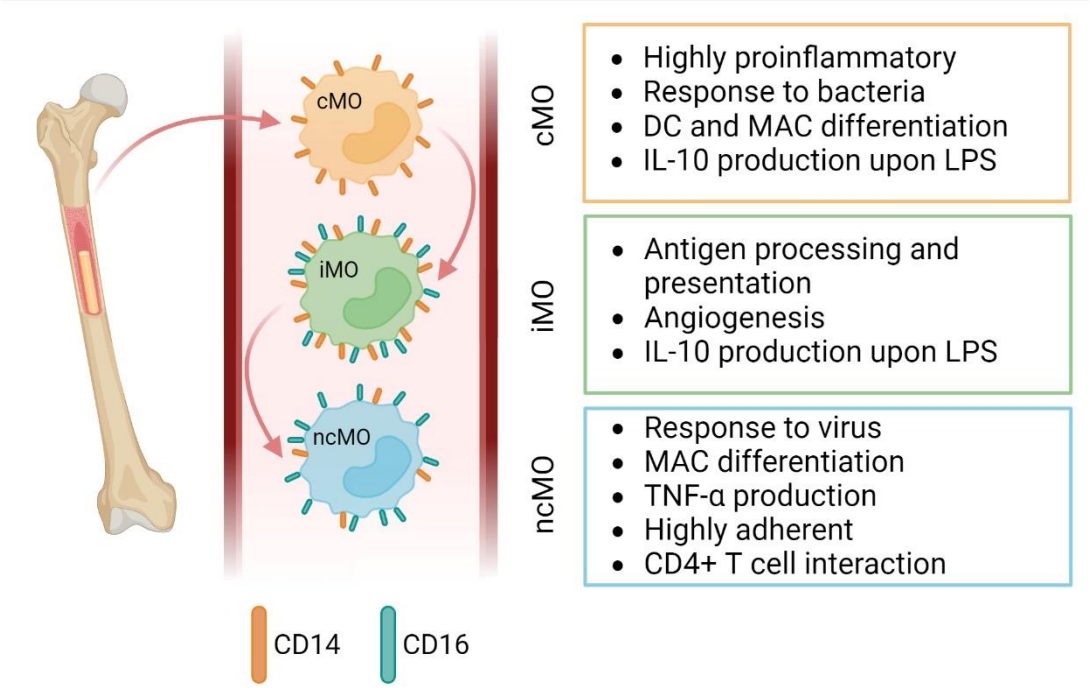


Figure 2. Monocyte subsets differentiation process. Scheme of the sequential differentiation process of monocyte subsets and some of the functions ascribed to each subset. cMO, classical monocyte; iMO, intermediate monocyte; ncMO, non-classical monocyte. Adapted from [29].

They are particularly effective producers of the cytokine interferon-α in response to viral stimuli, which is why they have a prominent role in the immune response to these pathogens [19]. The absence of expression of TLR1, TLR2 and TLR4 by ncMO portrays them as a deficient candidate for responses to bacterial pathogens which, as mentioned earlier, is mostly attributed to cMOs [21]. However, ncMOs are competent to respond to both double-stranded and single-stranded nucleic acid structures

through TLR3, TLR7 and TLR8 [19,21]. This ability might position them at the initiation of autoimmune diseases upon sensing human cells-released nucleic acids, as it is described for some forms of dendritic cells [30]. In fact, the presence of nucleic acids and immunoglobulins in the serum of patients of the autoimmune disease of Systemic Lupus Erythematosus induced the production of the inflammatory cytokines TNF-α and CCL3 by ncMOs [21]. In effect, they are established as the principal producers of TNF-α out of the monocyte subpopulations [21,22,31,32].

Despite all these pro-inflammatory evidences in favor of ncMOs, they have recurrently been described as the anti-inflammatory subset of monocytes. Largely due to their limited ability to respond to bacterial cues but also on account of their role as “patrolling monocytes”. Explained by their higher capacity of adherence to the endothelium due to their surface expression of CX3CR1 [33], ncMOs act as “intravascular housekeepers”: removing cellular debris and scavenging micro particles [34,35].

Last but not least, ncMOs have a significant role in the presentation of engulfed antigens to T cells, as seen in the spleen of mice [36,37]. In humans, they present an increased capacity to stimulate CD4+ T cells compared to the other monocyte subsets [38]. In physiological conditions, ncMOs induce a Th2 skewed differentiation [39], which is associated with the production of anti-inflammatory cytokines but also to immune response towards helminth pathogens and to some forms of chronic inflammation such as asthma or allergies [40].

Interestingly, recent work has shown that single cells forming the different subsets admittedly respond in an individual manner at the cellular level. Their classification into three groups simplifies their study, but single-cell technologies allow for detection of unique responses in individual cells. Monocytes appear to carry a distinct cell-inbuilt capacity to respond to certain stimuli [22], which further complicates their analysis.

The expected proportions of monocyte subsets might be influenced by multiple environmental factors, both physiological and pathological. In this way, age is positively correlated with expansion of iMOs and ncMOs [41,42]; female gender is associated with reduced levels of ncMOs [43]; a subset which increases up to three fold following exercise [43]. Furthermore, many pathological conditions can also affect monocyte proportions. In particular, CD16+ subsets are increased in a wide variety of disorders including Crohn's disease [44], stroke patients [45], colorectal cancer [46], obesity [47], tuberculosis [48] or arthritis [49]. These reports suggest a differential implication of the monocyte subsets in these diseases that might be caused by a unique role of each of them in every condition.

1.4 Monocytes and inflammation

Some of monocyte's major features include rapid mobilization and high plasticity capacity. Following inflammatory signals, they swiftly migrate from circulation in the peripheral blood to tissues, differentiating into a variety of cells that perform either proinflammatory or resolving functions [6].

As promoters of inflammation, the monocytes recruited during an inflammatory episode are more impactful in inflammatory molecules production and less in anti-inflammatory molecules production than their counterparts mobilized prior to the inflammatory episode [50]. These strongly inflammatory effector monocytes are specifically recruited in response to infections [51–54]. In these scenarios, recruited monocytes not only contribute to the inflammatory milieu by producing inflammatory molecules, but they also are crucial for the killing and clearing of the pathogens, functions for which the resident macrophages are insufficient. In some circumstances, the deletion of the recruited monocytes makes the model resistant to the external inflammatory induction as is the case for a mice model for multiple sclerosis [55,56].

Nonetheless, for a healthy homeostasis, the proper resolution of an inflammatory process is equally as important as the promotion of it. In this resolving context,

monocytes also play an active purpose. Several resolutive functions such as promotion of angiogenesis, clearance of neutrophils, clearance of cell debris, degradation of extracellular matrix to accelerate scar resolution as well as production of anti-inflammatory cytokines, they have all been ascribed to mobilized monocytes [57–60].

Dysregulations of either promotion or resolution of inflammatory processes might eventually occur, causing the emergence of pathogenic conditions. On the one hand, the inability to properly detect and eliminate pathogenic entities can lead to serious infectious processes such as sepsis, recurrent infections or cancer. On the other hand, the inability to extinguish immune reactions when the danger is exterminated can also lead to very serious situations as is the case with the severe responses to COVID-19 infections or autoimmune diseases.

1.5 Autoimmune disease

The term autoimmune disease comprises a wide range of pathologies where the checkpoints for self-identification and tolerance within the immune system have been bypassed. These defects result in the immune system attacking benign components of the self-organism. In these circumstances, given that the signal triggering the immune response is generated within the host, it can never be fully eliminated. This drawback leads to a constant state of inflammation and activation of the immune system that most certainly results in permanent damage to otherwise healthy tissues. Autoimmune diseases can affect individuals of any age having a higher prevalence among women than men. In general terms, their affectation can be restricted to a single organ or be generalized, which classifies them as organ-specific or systemic, respectively.

Although the term autoimmune disease encompasses a broad spectrum of disorders and manifestations, certain characteristics unite these pathologies. These premises are gathered in the Witebsky's postulates [61] which include: presence of autoreactive lymphocytes in the patients' blood, induction of the disease in animals

by transfer of these autoreactive lymphocytes or demonstration of autoreactivity at the tissue damaged, among others. In clinical everyday practice, the identification of antibodies recognizing self-components, or autoantibodies, is the most common approach for the diagnostic and classification of autoimmune diseases. Even though autoantibodies are a marker of autoimmune diseases, they are not necessarily the sole mediators of clinical manifestations and they might not be able to induce pathology by themselves.

Disease etiology is generally complex, involving a compendium of both genetic susceptibility and environmental triggers. It is not uncommon to find autoimmune tendencies in the same family, studies suggest a heritable trend for autoimmunity rather than for a specific disease [62]. For instance, one member of the family might suffer from systemic lupus erythematosus while another has a diagnosis for celiac disease. Genetically, MHC loci appear associated with susceptibility to a large variety of autoimmune diseases [63]. These molecules are responsible for the presentation of antigens to T cells, triggering their activation and differentiation. In the human species, MHC family of genes has a high number of polymorphism naturally. This variability confers the population overall protection against all antigens. However, it also makes some individuals more prone to autoreactivity [64].

It is established that environmental triggers are required to transform genetic susceptibility into disease incidence. The most recognized environmental trigger for autoimmune disease is the role of microorganisms and infections. Infections might trigger autoimmunity either by molecular mimicry existent between microorganism components and self-components or through non-specific mechanisms, by general immune system activation. For some pathogens, a combinatory effect of both mechanisms might help in the appearance of autoimmune. This is the case for Epstein-Barr virus infections, whose proteins are structurally similar to some human proteins and its infections can lead to an aberrant production of autoantibodies [65]. Smoke, drug use and treatment for other diseases have also been defined as triggers for autoimmunity, although the mechanisms are uncertain [66–68].

Even though the immune system is deregulated integrally in these situations, it is more perceptible in the adaptive immune system where lymphocytes specifically and aberrantly recognize self-components. Considering this, the adaptive immune system has been the research focus in the fields of autoimmune diseases historically. This thesis, however, focuses on the role of the less-studied innate immune system cells, particularly in monocytes.

1.5.1 Systemic Lupus Erythematosus

For this thesis project, autoimmunity research was approached through the study of the extremely complex systemic autoimmune disease named Systemic Lupus Erythematosus, hereafter SLE. It holds a varying prevalence across different populations that ranges between 16 and 110 affected individuals per 100000 persons [69]. It drastically affects more women than men with a ratio of 9:1 and it prevails in African American women, Hispanic women and women of other ethnic minorities [70]. Although some common alterations are shared among most patients, SLE is in reality a term encompassing a wide range of clinical manifestations. This variability hinders both its diagnosis and its posterior management [71]. Clinical manifestations can range from mild fatigue and intermittent joint pain to life-threating organ damage. SLE usually courses in relapsing-remitting episodes. Considering this, pharmaceutical management aims include long-term survival, prevention of flares and organ damage as well as improving health-related quality of life [72]. For an optimal monitoring of disease progression, a disease activity index was developed in 1985 [73] and updated in 2002 [74]. It is denominated Systemic Lupus Erythematosus Disease Activity Index or SLEDAI. It uses both clinical and biological parameters to formulate a score that ranges from 0 to 105. However, scores rarely exceed 20 and a score of 6 is often considered a threshold for active disease. This index has been successfully used in both clinical [75,76] and research [77,78] contexts.

Biologically, SLE's immunological hallmark consists in the presence in serum of antibodies that recognize nuclear components such as nucleic acids (DNA or RNA), proteins and complexes of DNA or RNA with proteins [79]. These antibodies, when

recognized by complement pathway components, produce immune complexes that deposit at both joints and vital organs. Consequently, callings for inflammatory responses are activated and result in chronic inflammation and tissue damage. Contradictorily, although the updated SLE classification criteria from 2019 requires positive anti-nuclear antibodies for the diagnosis of SLE [80,81], up to 30% of SLE patients enrolled in clinical trials are negative for these antibodies [82,83]. This fact accurately highlights the complexity and variability presented among SLE patients.

In SLE, the dysregulation of numerous cell types and components leads to a complex altered immune response. For instance, B cells are doubtlessly hyperactivated and bypassed of the control checkpoints for elimination of self-recognizing clones. Likewise, phagocytic cells such as macrophages are also altered. In SLE, cell death from PBMCs and cutaneous lesions is exacerbated. This factor tandems with a deficient phagocytic capacity of macrophages leading to an accumulation of cellular debris and release of self-antigens [84]. Importantly, this is not the only contribution of the non-functional myeloid compartment in SLE pathogenesis. Monocyte-to-macrophage differentiation also presents an imbalance ensuing a polarization towards proinflammatory M1 macrophages [85]. These macrophages inhibit cell proliferation and promote tissue damage, in contraposition to M2 macrophages that facilitate cell proliferation and tissue repair [86].

In relation to monocyte subsets, several works have studied monocyte subset proportions in SLE where highly varied results have been observed. Some indicate a higher percentage of cMOs [87], others show an increase in iMOs [88,89] and there are also studies reporting that ncMOs have increased proportions [90,91]. Finally, a single group also reported an absence of differences in the proportions of monocyte subsets in SLE in comparison to controls [92].

1.5.2 Interferon signaling

The production of cytokines is also altered in SLE. Cytokines are small soluble proteins that serve as mechanisms for communication and regulation among cells of the immune system. They can be secreted by various cell types but the myeloid

compartment is one of their major producers [86]. In SLE, there is an undeniable dysregulation of interferon (IFN) production and response from the cells of the immune system. In fact, this family of cytokines plays a fundamental role in the disease initiation and perpetuation [93]. This was first described when individuals receiving IFN-α therapy for unrelated pathologies developed an idiopathic SLE-like syndrome that improved and resolved upon treatment withdrawal [94,95].

The IFN family of cytokines is composed by three types of IFN, namely type I (mostly represented by IFN-α and IFN-β), type II (IFN-γ) and a less known type III (IFN-λ). In their canonical response pathways, type I IFN are recognized by a heterodimer of receptors formed by IFNα receptor 1 (IFNAR1) and IFNAR2 which are associated and activate the kinases known as Janus kinase 1 (JAK1) and Tyrosine kinase 2 (TYK2). These kinases phosphorylate the Signal Transducer and Activator of Transcription 1 (STAT1) and STAT2 which then dimerize and translocate to the nucleus. Either before or after entering the nucleus, they associate with IFN-regulator factor 9 (IRF9) and bind to specific DNA sequences known as IFN-stimulated response elements (ISRE) [96,97]. Type II IFN, on the other hand, signals through IFNGR1 and IFNGR2 that interact with JAK1 and JAK2. Upon activation, these kinases phosphorylate only STAT1 that homodimerizes and translocates to the nucleus. There, it binds to its target sequences containing the gamma-IFN activate site (GAS) [98]. Although type III IFN is recognized by other receptors (IFNLR1 and IL-10R2), its downstream pathway uses the same components and the same target genes as type I IFN, which challenges their discrimination [98] (Figure 3). However, cell-specific sensitivities differ between the two.

Under physiological conditions, IFN are released upon activation of TLRs by nucleic acids, bacterial components, viral proteins and other foreign antigens. Type I and III IFNs can be produced by most mammal cells in response to these stimuli. Type II IFN, however, is mostly produced by activated T and NK lymphocytes. Similarly, while type I IFN receptors are found in most cells of the immune system,

type II IFN exerts its functions primarily in the myeloid compartment [99]. Finally, type III IFN can only operate on epithelial and endothelial barriers [100].

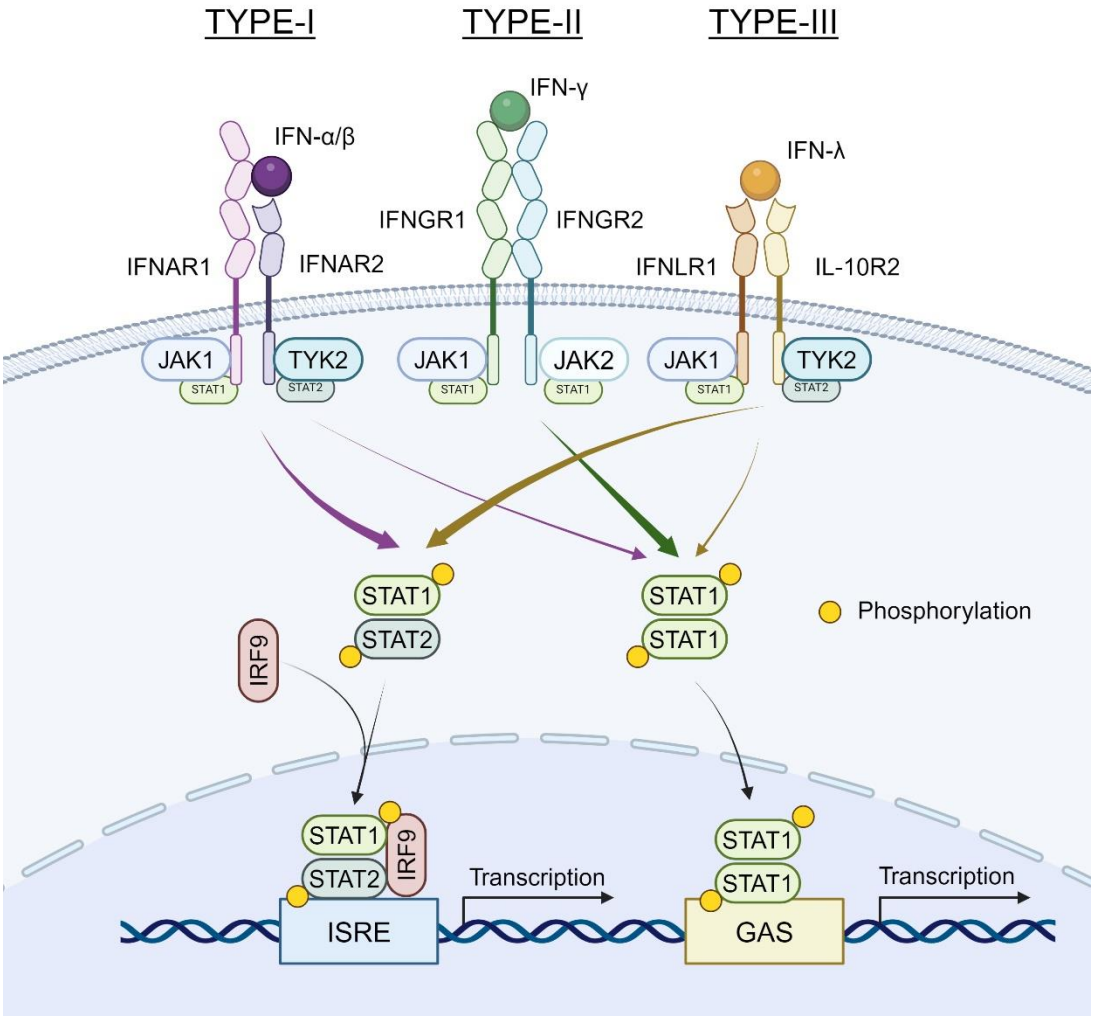


Figure 3. Canonical interferon signaling pathways. Upon interaction with their receptors, interferons start their downstream pathways by activating kinases, JAK1 and TYK2 in the case of Type I and III IFN, and JAK1 and JAK2 in the case of Type II. These phosphorylate STAT1 and 2 molecules which in turn form either heterodimers or homodimers and translocate to the nucleus. There they activate different target genes by recognizing either ISRE or GAS consensus sequences, respectively. Data from [96–98].

SLE has been mostly associated with type I IFN signature, however, as previously mentioned, several downstream components are shared between IFN cytokines. Thus, it is possible that other members of the IFN family are also deregulated in this pathology under the described type I IFN dysregulation. In fact, IFN-γ serum levels are elevated in patients' serum [101,102] and IFN-γ transcriptomic signature has also been observed in patients' monocytes [103]. Interferon secretion in SLE can be exacerbated in response of UV radiation, cell death or infections. Its effect on cells of the immune system are varied across the multiple compartments. IFN includes activation of myeloid cells to be potent antigen-presenting cells. It also promotes the activation and differentiation of CD4+ and CD8+ T cells towards an increased differentiation of Th17 cells and a suppression of Treg cells which favors the expansion of autoreactive T cells. Moreover, in B cells, activation and maturation are increased in response of IFN, as well as an increased antibody production via the effect activated monocytes have on B cells (reviewed in Saulescu et al [104]).

1.5.3 SLE etiology

As previously mentioned, SLE is such a complex disease that its epidemiology cannot be uniquely attributed to genetic modifications or to environmental triggers. Rather, it is a combination of both that leads first to its emergence and posteriorly to the development of the disease.

Indeed, SLE does have a genetic component, as it is specifically apparent in the cases affecting children. In these individuals, which represent 10-20% of lupus patients [105], exposure to accumulating triggering environmental impacts is limited. In effect, childhood-onset lupus is associated with high presence of SLE-associated polymorphism and higher genetic risk [106–108]. Through genome-wide association studies (GWAS), numerous single nucleotide polymorphisms (SNPs) have been associated with SLE susceptibility. They encompass more than 130 unique loci [109–111] suggesting a multiple-gene susceptibility to the disease. The affected loci include genes important for functions of both lymphoid and myeloid lineage cells. Several of

them involve genes of the IFN pathways such as *IFIH1, IRF7* or *IRF5*, as well as the MHC region such as *HLA-B* or *HLA-DQ*.

Nonetheless, environmental triggers withstand a fundamental part in the development of SLE, as evidenced by the discordance for lupus affectation in monozygotic twins [112–114]. In this regard, smoking [115], alcohol consumption [116], air pollution [117], obesity [118], diet [119], infections [120] or pesticides [121] are some of the currently accepted environmental triggers for SLE and other rheumatic diseases.

One of the principal mechanisms for cells to respond to environmental influences is through changes in gene expression. Gene expression regulation comprises an intricate combination of regulatory mechanism which entail the multiple steps required for a genome-contained gene sequence to become a functional protein product. These regulatory mechanisms cover the epigenetic, transcriptional, post-transcriptional, translational and post-translational levels. In this thesis, the gene expression regulation study has centered in the first level, the epigenetics.

1.6 Epigenetic regulation

Epigenetic regulation comprises a set of stable chemical modifications that cells use to regulate gene expression, specifically genetic transcription. Initially, they were described as covalent modifications on the DNA and the histones, the proteins that package it. Currently, the term has expanded to include any process affecting gene expression without alterations to the DNA sequence, such as non-coding RNAs. The main epigenetic mechanisms include methylation of cytosines in the DNA as well as histone modifications such as lysine and arginine methylation, lysine acetylation and serine and threonine phosphorylation of histones. In this thesis, the primary focus of epigenetic regulatory processes related to monocyte subsets and SLE is placed on DNA methylation. However, the entire set of regulatory processes collaborate to either facilitate or hinder the binding of both transcription factors and transcription machinery in order to conduct gene transcription.

1.6.1 DNA methylation

DNA methylation is the most studied form of epigenetic regulation. It consists in the addition of a methyl group to a nucleotide base. In particular, the addition affects a cytosine typically preceding a guanine, to which it is linked by a phosphate group. This dinucleotide is generally referred to as CpG. In some rare instances, the cytosine might be followed by another nucleotide base, which renames the dinucleotide as CpH, where H can be either adenine, timine or another cytosine. DNA methylation is critical for crucial cellular processes in mammalian development such as X chromosome inactivation and gene imprinting during embryonic development.

X chromosome inactivation, for example, happens at the early stages of embryonic development, in mammals it takes place only in female individuals. It is a mechanism aimed to equilibrate the gene dosage between the two genders, considering that females carry two copies of this chromosome, contrary to males. Through this process, each cell randomly selects one of its X chromosome copies to be transcriptionally silenced. With this end, permissive histone marks are removed, repressive ones are incorporated and CpGs are completely hypermethylated. This collection of epigenetic modifications causes an extreme change in the chromosomic tridimensional structure that avoids transcriptional expression of the majority of its genes. The heterochromatic state obtained is carried on through posterior cell divisions in physiological conditions [122,123].

In general, at least 70-80% of CpG dinucleotides in the mammalian genomes are regularly methylated [124]. Notably, there are genomic regions where CpG dinucleotides accumulate which are partially refractory to methylation. These regions are termed CpG islands and are typically found in the promoters of genes, often physically enveloping transcription start sites. It is estimated that around 70% of mammalian genes contain CpG islands proximal to their transcription start sites [125].

DNA methylation can carry opposing roles, inhibitory or permissive, depending on the genomic region where it occurs (Figure 4).

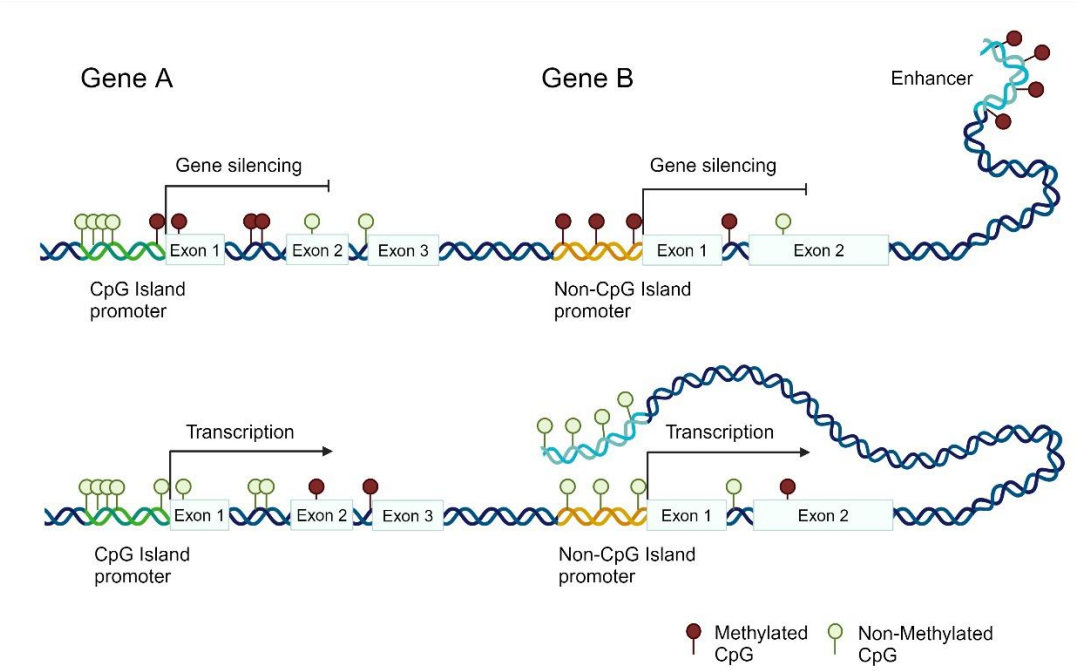


Figure 4. Effects of DNA methylation on gene expression. DNA methylation has different effects depending on the genomic region where it occurs. Methylation in promoter regions and enhancers usually associates with gene silencing. Methylation in the gene body has, generally, the opposite effect.

For instance, methylation in promoters is typically associated with loss in gene expression, particularly in promoters with lower densities of CpGs given that promoters with high CpG density repeatedly escape DNA methylation [126]. Similarly, DNA methylation of the first intron of a gene inversely correlates with the gene's expression [127], however, gene body methylation is positively correlated with gene expression [128]. Additionally, methylation of exonic regions might as well impact the product of alternative splicing [129]. Finally, enhancers are also affected by their methylation state. Methylation at enhancers has been associated with loss of gene expression while active enhancers show high levels of active demethylation [130].

Importantly, it appears that DNA methylation effect at promoters and enhancers is highly dependent on the sequence context. In this way, the binding of some transcription factors and transcription machinery might be differentially affected by sequence methylation.

The addition and removal of methyl groups in the DNA requires the participation of specialized enzymes:

- The addition of methyl groups in the DNA is catalyzed by DNA methyl-transferases (DNMTs). This family of evolutionarily-conserved enzymes contains six members in the human genome but only three are considered canonical due to carrying DNMT catalytic activity. They are DNMT1, DNMT3A and DNMT3B. DNA methylation is a heritable trait in cell division ascribed to its symmetrical presence on both DNA strands [131]. This fact allows the post-replicative maintenance of DNA methylation patterns by DNMT1, which copies the DNA methylation status of the template strand onto the newly synthetized strand [132,133]. A deficient activity of DNMT1 can lead to a passive demethylation process, by which methylation profiles are diluted through several rounds of cell replication. On the other hand, DNMT3A and DNMT3B are responsible for *de novo* methylation [134]. That is, the addition of methyl groups to specific previously unmethylated CpG sites.
- The active removal of DNA methylation is accomplished through a succession of reactions targeting the methyl group. In particular, three consecutive oxidation reactions facilitated by ten-eleven translocation (TET) enzymes [135]. With this process, the modified base shifts from a methyl-cytosine with a methyl group in the fifth position of the cytosine (5mC) to a 5-carboxycytosine (5caC). This modified nucleotide is then recognized by the base excision repair machinery that corrects small DNA lesions and substitutes the entire base with a non-modified cytosine [136]. TET family of enzymes contains three members in the human genome, TET1, TET2 and TET3. In human monocytes, TET2 is the most expressed member and its

proper expression is essential for DNA demethylation in this cell type [137]. Importantly, TET2 is the only member of the family that lacks the DNA binding domain. Consequently, it always requires recruitment to the DNA sequence by specific transcription factors [138]. Notably, oxygen is a cofactor for TET activity, thus oxygen absence in hypoxic conditions can modulate the activity of these enzymes, impacting DNA methylation [139].

The study of DNA methylation has advanced considerably in the last decades due to the development of sodium bisulfite modification of DNA technique in the early 90's [140,141]. This protocol is based on the differential chemical reaction of sodium bisulfite with cytosine and 5-methylcytosine bases. Through a sequence of chemical reactions, non-methylated cytosines become uracil while methylated cytosines are maintained as regular cytosines, although now unmethylated. This modification is strand specific. As a result, it is possible to perform either sequencing or differential probe binding to detect the initial methylation status of each individual cytosine in the sequence [142]. The combination of bisulfite modification with high-throughput sequencing techniques, either by whole genomic sequencing or using probe-based arrays, has enabled the simultaneous interrogation of DNA methylation status for the whole genome. This tandem has significantly advanced the field of DNA methylation study in the recent years.

As a consequence, the mechanisms by which DNA methylation regulates gene expression are now better understood. For instance, DNA methylation has an impact on transcription factors whose binding ability to their consensus sequence is strongly affected by its methylation status [143,144]. Not only that, but some transcription factors are now known to bind to densely condensed chromatin and induce the demethylation of their binding sites. These transcription factors have been termed pioneer transcription factors [145].

1.6.2 DNA methylation in SLE

In SLE, DNA methylation of several constituents of PBMCs has shown promising results or a better understanding of the disease. In particular, DNA methylation associates with disease susceptibility, classification and prognosis, disease activity as well as with response to treatment (Figure 5, reviewed by Ferreté-Bonastre et al [146]). Overall, SLE samples are characterized by a generalized hypomethylation pattern and a particularly altered epigenetic profile in the IFN signature genes.

DNA methylation and SLE susceptibility are intertwined at multiple levels. Methylation at susceptibility-associated SNPs presents different patterns in SLE patients and controls, which associates with disease development [147,148]. Additionally, DNA methylation might offer a potential explanation for the disparity in the incidence between males and females. Given that X chromosome inactivation is strongly linked to DNA methylation and that it is a chromosome highly enriched in genes involved in immune pathways, dysregulation of these genes' dosages might result in severe consequences [149]. Indeed, men require higher genetic risk / DNA methylation ratios to be as susceptible to the disease as women [150]. Concordantly, CD4+ T cells from women with SLE present greater hypomethylation and overexpression than their male counterparts for genes located in the X chromosome regulated by DNA methylation [151,152].

DNA methylation alterations are related with different clinical aspects of SLE and therefore can potentially be used for disease classification and prognosis. For example, different alterations of the DNA methylation profiles in various immune cell types are found between patients with and without renal affectation [153–155]. Similarly, DNA methylation relates to the classification of skin affectation by SLE, i.e., discoid rash and malar rash patients present differential epigenetic profiles on CD4+ T cells [156]. DNA methylation can further be used to classify patients according to disease severity into groups spanning from mild to highly severe, as shown in 2019 by Lanata and colleagues [157].

Regarding SLE disease activity, there are contradictory results regarding its correlation with DNA methylation. Several studies have found a strong correlation between activity (measured by SLEDAI) and DNA methylation at different cell types. Both at genes associated with interferon signaling pathway [78,158] and at genes related to other cytokine regulatory pathways [159,160]. However, other studies have found no association between DNA methylation profiles and disease activity [77,161,162]. The discrepancy in the results might illustrate the different ranges of disease activity included in the studies, or the inconsistency between using the patient's medical recent history or exclusively the activity score at the time of sample extraction.

Finally, DNA methylation can be also linked to response to treatments in SLE. Truthfully, medication itself can be responsible for changes in the epigenetic profiles. Glucocorticoids, a first-line treatment for SLE, is associated with higher levels of DNA methylation in PBMCs [163]. Also, the use of mycophenolate, an immunosuppressant used to manage kidney disease in SLE, modified the epigenome of patients at different regions. It associated with increased methylation levels in *IFNG* gene in CD4+ T cells, but not with global methylation changes [164]. Mycophenolate also affected the promoter region of *CD40L* gene targeting several histone modifications [165]. Additionally, although targeting DNA methylation has not been tested as a treatment for SLE, modifying the supplement of DNMTs' cofactor and methyl group donor S-Adenosylmethionine (SAM) showed promising results. In an initial study with inflammation-induced colon cancer, supplementation of SAM improved outcome and modulated inflammatory pathways [166]. Interestingly, restriction of SAM-containing nutrients such as folate and methionine in SLE could revert *in vitro* lupus-like gene expression in T cells from older donors [167].

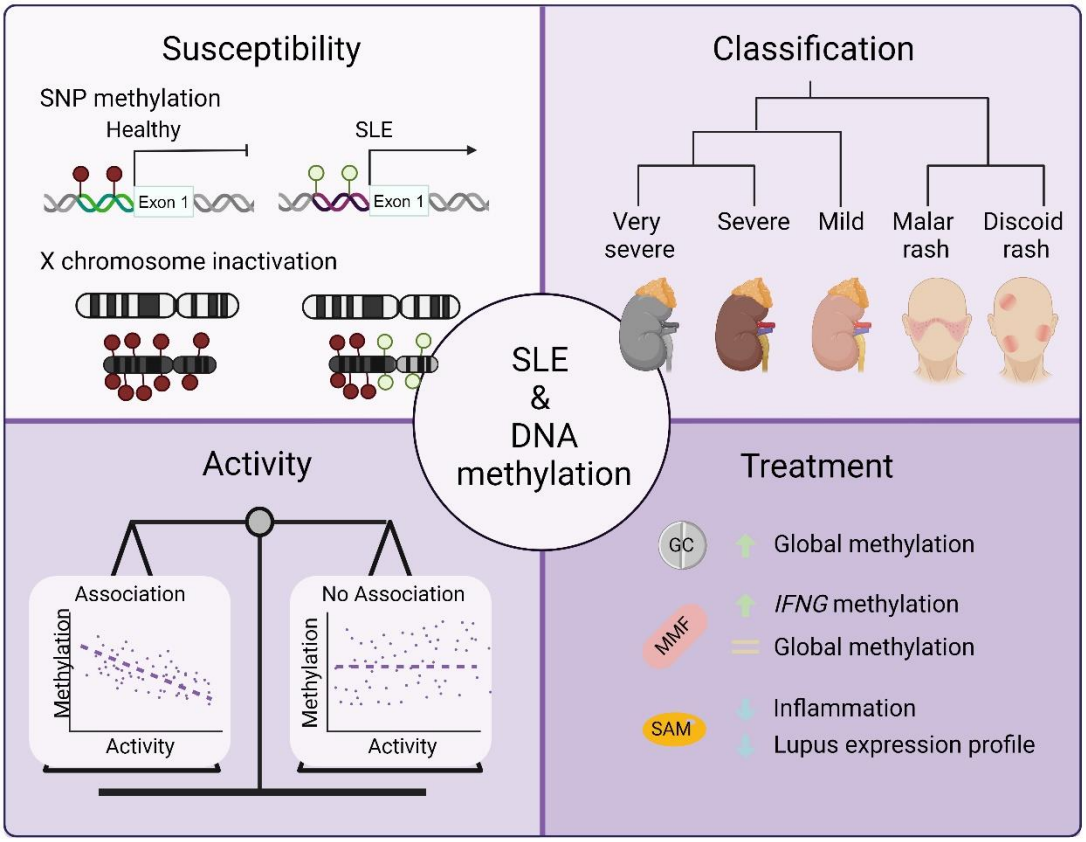


Figure 5. DNA methylation and SLE. Graphical abstract summarizing the results associating DNA methylation with SLE grouped in four categories. Susceptibility is affected by DNA methylation at SLE-associated SNPs and by increased demethylation of silenced X chromosomes in females. Classifications allows the categorization of systemic and cutaneous lesions of SLE based on DNA methylation profiles. Conflicting results regarding association of the methylome with disease activity represented by an equilibrated balance. Treatment might affect DNA methylation profiles as described for glucocorticoids (GC) and mycophenolate (MMF). Also, treatments targeting the reduction in DNA methylation (represented as SAM supplementation) might pose interesting benefits.

Altogether, DNA methylation in SLE has multiple relationships with disease development. As is usual for research in autoimmunity, most studies in SLE have mainly focused on the lymphoid lineage, particularly in CD4+ T cells. Dr. Bruce Richardson pioneered the study of DNA methylation in SLE [168]. His initial studies showed that inhibiting DNA methylation *in vitro* in CD4+ T cells induced autoreactivity in these cells [169,170], but not in CD8+ T cells[171]. Similar outcomes were observed on

patients receiving these treatments [172]. Understandably, the majority of studies for DNA methylation in SLE centered in CD4+ T cells. However, the long-forgotten innate immune system has been attracting attention in the field of SLE in recent years [173].

In particular, for DNA methylation studies in immune-mediated or inflammatory pathogenic conditions like SLE, the myeloid compartment and, in particular, monocytes are very interesting for multiple reasons. Firstly, monocytes are highly plastic and their phenotype and ability to rapidly adapt their state to environmental changes occurs through epigenetic modifications. Secondly, the fact that they are short-lived cells places them as an optimal candidate to evaluate first responses to changing environments. Since DNA methylation is a stable modification that can accumulate over time, short-lived monocytes are less prone to gather confounding modifications from previous conditions. Finally, and most importantly, by being terminal cells and not dividing into daughter cells, monocytes present themselves as an exceptional model to study active DNA methylation changes. This means that changes that take place *de novo*. In this regard, a loss of methylation between two monocyte conditions is not due to a dilution through cell division, but rather to an active process of demethylation through TET enzymes.

2 HYPOTHESIS AND OBJECTIVES

Monocytes are extremely plastic cells able to rapidly respond to a wide range of environmental changes. These responses or adaptations are reflected by changes in their epigenomes, which directly regulate or associate with gene expression changes, as shown by several studies from our group. Despite monocytes are often analyzed as a single cell type, they actually comprise a heterogeneous group of cells. Indeed, monocytes include various subsets or states with distinct functions and marked phenotypes throughout a differentiation process. There is evidence suggesting a diverse implication of these cell groups in pathology. In the present thesis, we aim to better understand the differences in monocytes subsets and their potential roles in disease.

With these considerations in mind, we divided the overall objectives in the following specific aims:

1. To characterize the epigenetic and transcriptomic differences that delineate the different monocyte subsets phenotypes and elucidate their roles in monocyte differentiation.

Monocyte subsets, namely cMOs, iMOs and ncMOs, arise from a sequential differentiation process. In humans, the mechanisms driving this differentiation process are poorly known. Through the study of epigenomic and transcriptomic profiles of steady-state monocytes subsets, we aimed to identify novel elements regulating their sequential differentiation and propose mechanisms that could be crucial to this differentiation process.

2. To describe the distinct phenotypes of monocytes subsets in pathology, particularly in Systemic Lupus Erythematosus, and elucidate their individual contributions to the disease.

Several groups have reported differences in the proportion in monocyte subsets in many pathological inflammatory conditions, including SLE. This fact strongly suggests an unequal contribution of monocyte subsets in these disorders. Through

the combination of epigenomics and transcriptomics analyses of samples corresponding to the three monocyte subsets from SLE patients, we aimed to infer their potential contribution to pathogenesis and propose novel mechanisms regulating them.

3 MATERIALS AND METHODS

For the development of this thesis, we have applied bulk multi-omics analyses of monocyte subsets from fresh blood samples from either healthy donors or from SLE patients. Monocyte subsets were obtained by flow cytometry sorting. Some validation experiments were performed *in vitro* with cells from buffy coats. Finally, we have also performed validatory analyses using public single-cell RNA-seq datasets. Details of the methodology are provided below.

3.1 Sample acquisition

For monocyte subsets separation from PBMCs, a flow cytometry sorting strategy similar to many published before was used [90,174,175]. Two sample cohorts were collected for the study of DNA methylation and transcriptome analysis, one in steady-state and the other including both pathological SLE samples and healthy controls. For the steady-state study, blood was obtained from seven healthy donors (HD) including both genres and an age range around 20-30 years of age.

For the SLE study, 20 patients with SLE and 13 HDs were included. The HDs were matched with SLE donors based on age and sex. 14 out of 20 SLE donors were of Caucasian ethnicity and, similarly, the majority of HDs belonged to the same ethnic group. All participants gave both oral and written consent for their blood to be used for research purposes. The study was approved by the Committee for Human Subjects of the Bellvitge Hospital ethics committee (PR275/17). All the patients with SLE fulfilled the 2019 European Alliance of Associations for Rheumatology/American College of Rheumatology classification criteria for SLE [80]. Samples from each patient with SLE were collected at two different time points, the first one at the onset of a new flare and the second one at the subsequent visit. For patients with SLE, the Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2k) [74] was registered at each extraction date. Definition of a flare episode was done with Safety of Estrogens in Lupus Erythematosus-SLEDAI Flare Index, which can be used with any version of SLEDAI [176,177].

For validation and *in vitro* assays, blood was obtained from anonymous buffy coats through the Catalan Blood and Tissue Bank, which follows the guidelines of the World Medical Association Declaration of Helsinki. All donors provided an informed consent before donating the first blood sample.

3.2 Monocyte subsets isolation by flow cytometry sorter

25 mL of whole blood were processed within 24 hours of collection by laying on Lymphocyte Separation Solution (Rafer, Zaragoza, Spain) and centrifuging without breaking in order to obtain PBMCs. The cells were washed with phosphate buffered saline (PBS, Gibco) and residual erythrocytes were lysed for five minutes with Ammonium-Chloride-Potassium (ACK) lysis buffer. PBMCs were either stained or cryopreserved in fetal bovine serum with 10% dimethyl sulfoxide. Cryopreservation was particularly used for the second cohort with the aim to gather several samples for flow cytometry sorting on the same batch. The day of cytometry sorting, samples were quickly thawed at 37°C.

Staining was performed in staining buffer (PBS containing 2mM of EDTA [Sigma-Aldrich] and 4% of Fetal Bovine Serum [FBS, Gibco]) with CD3-FITC, CD15-FITC, CD56-PE, CD16-APC, CD14-APC-Vio770 (Miltenyi Biotec) and CD19-FITC (BD Bioscience). Cells were incubated for 20 minutes on ice protected from light and posteriorly washed once with staining buffer. Finally, cells were resuspended in staining buffer with viability staining (DAPI).

Sorter gating strategy included the selection of single live cells by forward and side scatter intensities as well as absence of viability staining. Monocyte cells were selected by negative gating for CD3, CD15, CD19 and CD56 staining as recommended by [90,174,175], resulting in the acquisition of the monocytic fraction. This negative fraction was further analyzed for CD14 and CD16 cell surface markers for the separation of cMO (CD14+ CD16−), iMO (CD14+ CD16+) and ncMO ($CD14^{dim}$ CD16+). Purity of the fractions separated was always maintained above 98%.

3.3 Genomic DNA and RNA double extraction

Sorted samples were pelleted and frozen in RLT Buffer+10% β-mercaptoethanol as recommended by the kit Allprep DNA/RNA Micro, Mini (Qiagen). After collecting all the samples, double extraction of DNA and RNA was performed in as few batches as possible following manufacturer's instructions. DNA was quantified with Qubit™ dsDNA HS Assay Kit (Invitrogen) and RNA was quantified with Nanodrop.

3.4 DNA preparation

For DNA studies in steady-state, the three subsets from five healthy individuals were used. In total, 15 samples were modified and hybridized. For the second cohort, the three monocyte subset samples from 20 patients in the first and second visits as well as 10 samples from HDs were used; in total, 150 samples were analyzed. For all samples where possible, 250 ng of genomic DNA were modified with bisulfite with EZ DNA methylation Gold kit (Zymo Research) following manufacturer's instructions.

Modified material was hybridized in Infinium MethylationEPIC V.1.0 arrays, which interrogate around 850 000 CpG positions distributed throughout the genome, covering more than 99% of the reference sequence (RefSeq) genes. Fluorescence of the probes was read with BeadArray Reader (Illumina, Inc.) and image processing and intensity data extraction were performed as previously described [178]. Readout information output is retrieved as a beta value that ranges from 0 to 1 and represents the methylation level of a particular CpG. Moreover, a detection p-value for each position is also retrieved indicating the reliability of the measurement. Since these arrays only accommodate eight samples each, a sample distribution strategy was devised to mitigate potential batch effects. This was particularly important in the second cohort which contained more sample variability. With this aim, disease state, subset, donor's age and sex, visit number, SLEDAI activity, as well as the time the sample spent frozen, were distributed in a balanced manner across arrays [179].

3.5 RNA preparation

For RNA study on the first cohort, monocyte subsets sorted samples from three different individuals were included, in total 9 samples were sequenced. RNA-seq library preparation and sequencing were performed by Novogene (Cambridge). Due to low amount of starting material, Ultra RNA Library Prep Kit for Illumina® was used for library preparation. Samples were sequenced with 150-basepairs paired-end protocol with Illumina Novaseq PE150 platform. All samples passed quality control assessment. RNA sequencing data was annotated with HISAT2 software [180,181] and analyzed with DESeq2 v1.32.0 package [182].

In the SLE cohort, samples from the three monocyte subsets of 13 patients in the first visit, as well as seven samples from HDs were used. In total, 60 samples were sequenced in a 100-basepairs paired-end manner. In this case, libraries and sequencing were performed by BGI Genomics (Hong Kong) with low-input transcriptome sequencing Smart Seq-based method and DNBseq platform. Approximately 50 million reads were obtained per sample, and all samples passed sequencing quality control assessed with FastQC [183]. Samples were annotated and normalized with Kallisto [184] since it is a more optimal method in terms of time and memory usage. Given that 60 samples were consecutively processed, these assets were of the essence.

3.6 DNA methylation profiling analyses

Data from DNA methylation studies were analyzed following the pipeline described for the shinyÉpico package [185]. In the end, after removing unreliable probes with detection p-values above accepted threshold (p-value < 0.01), CpHs (where H = A, C or T), SNPs and X/Y chromosomes, samples were normalized using noob+quantile algorithms. Quality control was performed and sample composition was checked with the function *estimateCellCounts* from minfi package V.1.42 [186]; some samples were removed due to undesired scores in either of the tests. In brief, four samples from each subset in the first cohort and 137 samples in total for the second cohort were retained for further analysis.

For differential analysis, beta values were transformed to M values with lumi package V.2.48 [187]. This logarithmic transformation is crucial to reduce the variability in variance that is naturally present in beta values, a term known as heteroscedascticity. The linear model posteriorly applied to identify differentially methylated positions assumes that the variance of the data is constant, i.e., it assumes the data to be homoscedastic.

The limma [188] function *lmFit* was used for statistical analysis. In the generation of the linear model. *Arrayweights* argument was selected to weight the samples depending on their reproducibility. At this point, the analyses performed for the first and the second cohort were slightly different since they were responding to different questions:

- For the first study, three independent contrasts were performed (cMO-iMO, cMO-ncMO and iMO-ncMO). In all comparisons, samples from the groups compared originated from the same donor, for this reason the variable donor was selected as a covariable of the statistical model. *eBayes trend* and *robust* options were enabled in the computation of Bayes statistics. The objective with this is to further reduce the variability of the variance that is dependent on the signal intensity or on the methylation value. Differentially methylated positions (DMPs) were defined as those CpGs with a differential beta value > 0.2 and adjusted p-value < 0.05 (Bonferroni adjusted) in either of the comparisons.
- For the second cohort, since the sample group was larger and more information about potential covariables had been collected, we performed a preliminary analysis to identify environmental factors with a significant effect on methylation values. In particular, we analyzed the association of SLE samples with different types of medication, recent UV exposure, smoking and alcohol consumption habits, diet and prolonged exposure to environmental contaminants, stress as well as the use of hormonal contraceptives. The absence of a significant association led to the exclusion of these factors from

the analysis design. The only covariate clearly associated with differences in DNA methylation was the monocyte subset. Thus, we decided to perform the comparisons between HD and SLE samples individually for each subset. In this analysis, only samples from the first visit were used to avoid biological replicates. CpGs with a difference in beta value > 0.1 and an adjusted p-value < 0.2 were considered as DMPs. This more generous threshold was necessary in order to obtain enough DMPs for a significant posterior biological interpretation. Reproducibility and reliability of the results were validated by performing 100-fold permutations of the samples' distribution among the groups compared. In only 5% of these permutations the number of DMPs was higher than that obtained with the original groups. More than 40% of the permutations resulted in a total of 0 DMPs. These results support the specificity of our DMPs in the sense that the differences described are strongly associated to the two groups compared: HD vs SLE.

DMPs were grouped into clusters with the function *hclust* from the stats package with clustering algorithm parameter *method* set to average and a *distance function* using Pearsons's correlation.

Annotation of CpGs to their closest gene with information about genetic location was performed with the function annotatePeak from the package ChIPseeker V.1.32 [189,190] and the annotating dataset TxDb.Hsapiens.UCSC.hg19.knownGene. The annotation relative to CpG Islands was performed with the package annotatr v1.18.1 [191] and the annotation hg19_cpgs. In this dataset the categories include "CpG Islands", "shores" which are defined as 2 kb upstream/downstream of CpG islands, "shelves" that are 2 kb further from shores and the rest is considered "outside CGI".

Transcription factor motif enrichment analysis was used to study the prediction of individual transcription factors binding to the regions neighboring the DMPs. This analysis uses a database of TF consensus binding sequences to measure the enrichment of said sequences in the regions englobing DMPs. The study was performed with HOMER v4.11 [192] software with the function *findMotifsGenome.pl*,

using a window of either 250 or 500 base pairs around the DMP. The argument *–cpg* was used to normalize the CpG% content.

Enrichment analysis of chromatin functional states were performed with public data from CD14 primary cells from the Roadmap Epigenomics Project [193,194]. This data consists of information from 5 histone modification marks that constitute a model of 15 different chromatin states. Enrichment and significance of the DMPs in each of these states were estimated by Fisher's exact tests.

Gene ontology (GO) and pathway analysis of DMPs were performed with GREAT software package for R rGREAT v2.0.2 [195,196]. In this, the two closest genes were used to annotate CpGs to genomic regions with argument *rule* set to twoClosest, the version of GREAT was set to 4. GO terms were considered significant when adjusted p-value < 0.05. All annotated CpGs in the EPIC array were used as background in all these analyses.

Gene Set Enrichment Analyses (GSEA) were performed with mCSEA package v1.18 [197]. This analysis ranks the CpGs according to their methylation difference in the two compared groups. It then measures the enrichment of a list of positions in this rank. If the studied positions are equally distributed along the ranking of CpGs it results in no enrichment. However, studied positions can also be enriched in either side of the ranking indicating a general methylation status in either of the conditions. This analysis was only used in the first cohort to study the enrichment of IFN-response demethylated CpGs in the comparison cMO-ncMO. For this, the probes were ranked by their t-value in this comparison with the function *rankProbes*.

In the second study, correlation of methylation and SLE activity as measured by SLEDAI was performed with a Spearman's correlation with an estimate threshold of 0.7 and p-value < 0.001.

Plots were generated with ggplot2 [198], gplots V3.1, and gviz [199].

3.7 Bulk RNA sequencing analyses

After alignment to hg38 and normalization of the data, differentially expressed genes (DEGs) analyses were performed with DESeq2 v1.32.0 package [182]. In the second cohort, the sample labeled as SLE15 was excluded from posterior analyses due to having an outlier profile. For both studies, DEGS were defined as protein coding genes whose statistical analyses resulted in an adjusted p-value < 0.05 and Fold change (FC) < 0.5 or > 2. DEGs were found by *lfcShrink* function with Ashr algorithm. This function performs a shrinkage on the fold changes of the transcripts with low counts to correct for the higher tendency of showing large inexplicit fold changes in these. Variance Stabilizing Transformation (VST) values and normalized counts provided by DESeq2 were used for visualization purposes.

GO analyses were performed with *enrichGO* function from clusterProfiler package v4.6.2 [200] with org.Hs.eg.db database v3.16 [201]. Statistically significant and biologically relevant functions were selected for representation.

Inference of transcription factor activity by enriched regulon analysis was performed with VIPER 1.26.0 package [202]. Either DoRothEA [203] dataset A, B and C or Collection of Transcriptional Regulatory Interactions (CollecTRI) [204] were used as datasets for these analyses. The genes were ranked by their adjusted p-value combined with the sign of the FC. For the second study, a transcription factor's targets score was generated for STAT family members. These were calculated with the function *run_viper* from the DoRothEA package [203] and the CollecTRI database [204].

GSEA was performed with function *GSEA* from clusterProfiler [200] with the genes ranked by adjusted p-value and log2FC. This analysis works similarly to the one described previously for the DNA methylation section. In this case, it is the genes that are ranked between the two conditions according to a set of statistical parameters. The enrichment of a set of genes is analyzed along this rank of genes. Two groups of gene sets were used for these analyses. On the one hand, public gene sets

containing genes that increase their expression upon IFN stimulation. On the other hand, original gene sets from the previously defined DMPs were created by annotating the DMPs to their closest genes with annotatePeak from the ChIPseeker package [189,190].

Plots were generated with ggplot2 [198].

3.8 Cell culture

For primary monocyte cell culture, CD14+ monocytes were obtained from PBMCs from buffy coats by positive selection with magnetic CD14 MicroBeads (Miltenyi Biotec) following manufacturer's protocol. Monocytes were resuspended in Roswell Park Memorial Institute (RPMI) Medium 1640+ GlutaMAX™ (Gibco, ThermoFisher) and allowed to attach to the cell culture plate. After 15 minutes, the medium was changed with RPMI containing 10% heat-inactivated human serum (Merck), 100 units/ml penicillin/streptomycin (Gibco, ThermoFisher) and the proper stimulus. Stimulus consisted in 100ng/ml of IFN-γ or 50ng/ml of either IL-12, IL-15 or IL-21 (Peprotech). Samples were collected after 6h or 24h in culture.

3.9 RT-qPCR

Total RNA was extracted from treated samples with Maxwell RSC simplyRNA Cells kit (Promega) following manufacturer's instructions. 250ng of total RNA were reverse-transcribed to cDNA with Transcriptor First Strand cDNA Synthesis Kit (Roche). qRT-PCR was performed in technical triplicates for each sample using SYBR™ Green PCR Master Mix (Applied Biosystems), and 7.5 ng of cDNA per reaction. The primers used for *TBX21* sequencing were: AGGTGTCGGGGAAACTGAG (Forward) and ACCACGTCCACAAACATCCT (Reverse). The primers used for housekeeping gene *RPL38* were: TGGGTGAGAAAGGTCCTGGTC (Forward) and CGTCGGGCTGTGAGCAGGAA (Reverse). The average value from all technical replicate was obtained. Then, the standard double-delta Ct method was used to determine the relative quantities of

target genes, and values were normalized against the housekeeping gene. Plot was created with R package ggplot2 [198].

3.10 Western blotting

Proteins HIF-1α, T-bet, STAT1 and p-STAT1 were detected by Western blot technique. For this, sorted or treated samples were resuspended in Laemmli 1X loading buffer after two washes with PBS. DNA material from the samples was broken down by 20 seconds of ultrasonic processing at 60% amplitude to facilitate sample loading. Sample concentration estimation was measured by calculating the amount of DNA within the sample with Nanodrop and using it as an indicator of cellular concentration. Equivalent, boiled and 1:100 beta-mercaptoethanol treated samples were loaded into 12% SDS-PAGE acrylamide gels. Immunoblotting was done on polyvinylidene difluoride (PVDF) membranes following standard protocols. After blocking of the membranes with 5% Difco™ Skim Milk (BD Biosciences) they were blotted overnight with primary antibodies. Three consecutive 10-minute washes were performed with TBS-T (50 mM Tris, 150 mM NaCl, 0.1% Tween-20) followed by 1-hour incubation with HRP-conjugated mouse or rabbit secondary antibody solutions (Thermo Fisher) diluted in 5% milk in a dilution of 1:10000. Finally, proteins were detected by chemiluminescence using WesternBright™ ECL (Advansta). The antibodies and the dilutions used included: Anti-Tbet 4B10 (Invitrogen, 1:500), Anti-STAT1 42H3 (CellSignaling, 1:1000), Anti-phosphorylated (Tyr701) STAT1 58D6 (Cell Signaling, 1:1000), Anti-HIF-1α #3716 (CellSignaling, 1:500), Anti-H3 ab1791 (Abcam, 1:20000) and Anti-GAPDH 14C10 (CellSignaling, 1:10000).

3.11 Single-cell RNA-sequencing analysis and integration

Each study was integrated and expanded with the analysis of a public dataset of single-cell RNA-sequencing.

- For the first study, an object generated from the PBMCs of an individual with a homozygous loss-of-function mutation in the gene encoding the transcription factor T-bet, along with his heterozygous male progenitor was

used (EGAS00001004504 [205]). Data were loaded with the function *CreateSeuratObject* from the Seurat package v4.3.0.1 [206]. Good quality cells were considered as those with a percentage of mitochondrial genes < 20%, and a number of features < 2500 and > 200. After filtering, myeloid cells were not re-integrated, the object of 4173 cells was used for clusterization and annotation.

- For the second study conforming this thesis, we used the dataset GSE174188 [207]. It contains data from PBMCs from 162 SLE patients and 98 HD. The dataset was initially in a Python processed format and we converted it to a Seurat object with the function *Convert* from SeuratDisk. Good-quality cells were considered as those with a percentage of mitochondrial genes <15%, a number of counts >1500 and a number of features <3500 and >500. Analyses were performed exclusively on the myeloid fraction of the total object. After filtering, myeloid cells were re-integrated (with the parameter *k.weight* set to 50), considering batch information. Doublets with T cells and platelets were removed. An object with 266k cells was obtained and used for clusterization and annotation. Given the fact that the ethnicities represented in the bulk RNA-seq dataset were primarily European and Hispanic, for the bulk integration analyses, these two cohorts were filtered resulting in a final object with 154k cells.

For Uniform Manifold Approximation and Projection (UMAP) representation, the functions *ScaleData, RunPCA, FindNeighbors, FindClusters* and *RunUMAP* were used sequentially, with the first 30 dimensions.

Cells from a single-cell object can be grouped together according to their transcriptomes similarities. This process is called clustering and it depends on the parameter resolution which decides the granularity of the process. Increased resolution leads to higher number of smaller clusters. The analysis of several cluster resolutions allows to better understand how cells resemble and group each other. A clustering tree is a representation of the clustering results at different resolutions and

how these groups of cells correlate along them. The clustering trees were calculated and plotted with clustree package v0.5 [208].

In order to annotate cell clusters to particular cell types, the top differential genes or markers for each cluster were analyzed. Top markers for each cluster were identified with the function *findAllMarkers* with default arguments. GO of top 50 markers was calculated with *enrichGO* function from clusterProfiler package v4.6.2 [200] using as background all the genes included in the analysis.

Communication between groups of cells can be inferred based on the expression of ligand and receptor molecules in the different groups of cells. In single-cell analyses, Cellchat package v1.6.1 [209] facilitates this process. This package was used for the inference of cell-cell communication between the clusters of the myeloid compartment and the rest of compartments of the PBMC object, particularly in the first study.

Pseudotime analysis in a single-cell object allows the computational placement of the cells from a continuous biological process in a pseudotemporal trajectory, based on their progressively evolving transcriptomes. Pseudotime analysis was only used in the second study of this thesis and its aim was simply to order the cells from an initial cMO state to a final ncMO one. Through this, we were able to better understand and represent the changes taking place along the differentiation process. Pseudotime was calculated with the package monocle3 [210–212]. The cells' feature displayed throughout the differentiation was the expression score of the STAT1 target genes that was calculated with the function *AddModuleScore* from Seurat package [206].

Integration between single-cell data and bulk data can also be acquired by estimation of single-cell cluster representation in the bulk data. With this end, the software CIBERSORTx [213] creates signature matrices from each cluster defined in a single-cell object. Later, the proportion of expression of these expression signatures is estimated on the bulk samples. Statistical inference of the proportions differences

was conducted using a linear model with the function *lm* and a two-sided Wilcoxon rank-sum test.

4 RESULTS

The Results section of this thesis has been divided in two parts, each addressing one of the thesis' objectives. In the first part, the differentiation process from cMO to ncMO in steady-state is studied through the analysis of epigenomics and transcriptomics of monocyte subsets samples from healthy donors. In the second part, the implications of monocyte subsets in the autoimmune disease SLE are assessed with a cohort that includes samples from both SLE patients and HDs. The study includes epigenomics and transcriptomics approaches, as well as an integration with a public single-cell transcriptomic sequencing dataset.

4.1 PART 1. Epigenomic and transcriptomic changes in monocyte subsets differentiation

4.1.1 DNA methylation changes associated to interferon-gamma during classical to non-classical monocyte differentiation

To investigate the regulatory mechanisms leading monocyte subsets differentiation, we initially purified by flow cytometry the three monocyte subsets (Figure 6A) from healthy donors following a negative gating strategy similar to those published previously [90,174,175] (Figure 6B). Bisulfite modified DNA obtained from the resulting sorted cells was used to analyze the DNA methylation profiles of the three subsets by Infinium MethylationEPIC V1.0 bead arrays. The analysis revealed a total of 2625 CpG positions differentially methylated between at least two of the subsets (False Discovery Rate [FDR] < 0.05 and differential of beta > 0.2) (Figure 6C). Hierarchical clustering of the methylation patterns at these positions enabled the discrimination of two distinct clusters. Cluster M1 comprises 1905 differentially methylated positions (DMPs) that experience DNA demethylation during the differentiation from cMO to ncMO. Cluster M2 consists of 720 DMPs that undergo gains in DNA methylation from cMO to ncMO. In both clusters, the iMO subset displays intermediate methylation levels between the other two subsets (Figure 6C).

Principal component analysis of the DMPs revealed a higher similarity between iMO and ncMO than with cMO (Figure 6D). CpGs from both clusters annotated

preferentially to regions outside CpG islands while annotation to CpG islands was clearly underrepresented in comparison to the background (Figure 6E, left). Concordantly, annotation to genomic regions showed a clear reduction in promoter regions representation in both clusters in favor of an increase in distal intergenic and intron regions (Figure 6E, right). Enrichment analysis of chromatin functional states showed that DMPs from both clusters were significantly enriched in enhancers as well as significantly underrepresented in transcription start sites (TSS) and repressed regions (Supplementary Figure 1A)

Gene ontology (GO) analyses revealed that CpGs from both clusters are associated to immune response functions such as leukocyte activation or regulation of immune response. CpGs from cluster M1 were repeatedly associated with myeloid activation functions. Notably, the ontology terms “response to interferon-gamma" and “positive regulation of interferon-gamma production” were significantly enriched with an adjusted P-value < 0.05 exclusively in the DMPs from cluster M1 (Figure 6F). Among the genes annotated by proximity to these DMPs we found some with a functional relevance in immunological context such as *FLT1*, *HCK*, *PRDM1* or *PTGS2* for cluster M1 (Supplementary Figure 1B). Also, CpGs from cluster M2 were enriched for “positive regulation of fractalkine biosynthetic process” function, which has been described to be relevant for reducing apoptosis levels in the different subsets of monocytes. In cMO particularly, it achieves this by reducing intracellular levels of reactive oxygen species (ROS) [214]. Some example genes annotated to these DMPs were *SPI1*, *TLE1*, *ITGB7* and *SBNO2* (Supplementary Figure 1B).

Transcription factor (TF) binding motif enrichment analysis of the regions surrounding the identified DMPs showed a clear representation of known myeloid lineage TFs. For example, cluster M1 presented an overrepresentation of factors such as PU.1 and KLF4 [215,216]. as well as several interferon regulatory factor family members. In cluster M2 we observed a redundant representation of CEBPs [215] as well as NFIL3 [217] have previously been described to play a crucial role in myeloid lineage and differentiation.

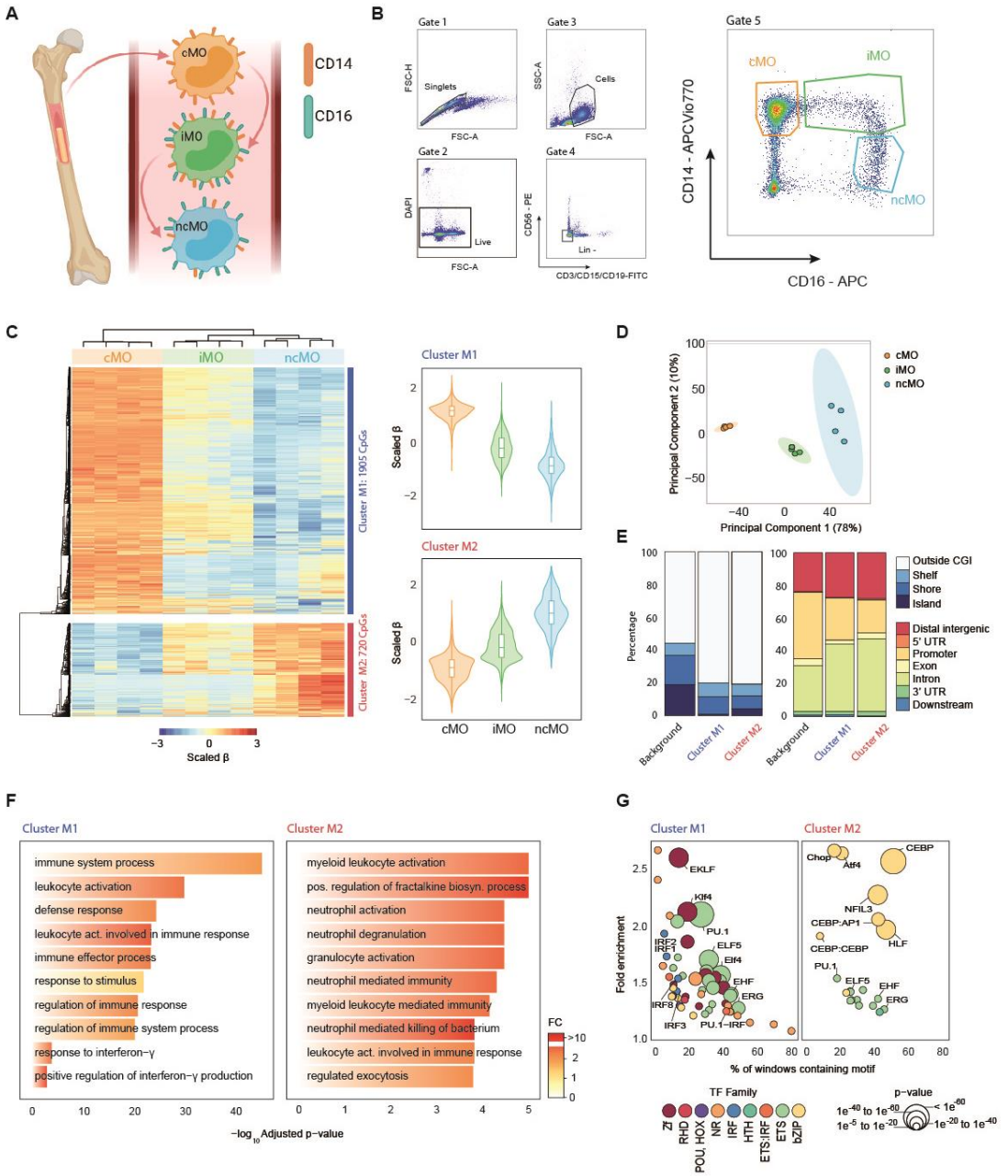


Figure 6. Epigenetic differences of monocyte subsets at steady-state. A. Graphical representation of the differentiation model of monocyte subsets in the bloodstream. cMO is classical monocytes, iMO is intermediate monocytes and ncMO non-classical monocytes. **B**. Flow-cytometry gating strategy for the acquisition of monocyte subsets. **C**. Heatmap plot and cluster analysis (left) of the differentially methylated positions (DMPs) between at least two monocyte subsets (FDR < 0.05, differential of beta > 0.2). The color annotation of the lateral bar represents the association with cluster M1 in blue (loss of methylation in the differentiation from cMO to ncMO) or cluster M2 in red (gain of methylation in the differentiation). Violin plots

(right) of the DMPs normalized methylation values separated by cluster association. **D**. Principal component analysis of the methylation values of the DMPs. First and second components are represented. **E**. Genomic annotation proportion of DMPs from each cluster in relation to CpG Islands (CGI, left) or gene related regions (right). **F**. Gene ontology terms associated with the DMPs in each cluster. Bar length represents log-transformed binomial q values of the term enrichment. Bar color correlates with the term's fold change (FC). For visualization purposes, color scale has been truncated after 10 but FC of "Positive regulation of fractalkine biosynthetic process" is 256 and that of "Neutrophil mediated killing of bacterium" is 43.38. **G**. TFs significantly enriched by the motifs discovery analysis by HOMER of the regions surrounding the DMPs from each cluster.

It is interesting to note that many of the TFs enriched in both clusters. However, different TFs appear to have major roles in the different subsets. Importantly, interferon regulatory factors (IRF) such as IRF1, IRF2, IRF3 and IRF8 were enriched only in cluster M1 indicating, again, a potential demethylation-associated activation of the pathways regulated by interferon in the positions associated with demethylation in ncMOs (Figure 6G). Together, these results reinforce the idea that the three monocyte subsets are both part of a continuous differentiation process and, at the same time, different enough to deserve individual characterization.

4.1.2 Transcriptomic reprogramming from classical to non-classical monocytes unveils antagonistic roles for HIF-1α and T-bet

We further characterized the three monocyte subsets by profiling their transcriptomes. RNA-sequencing analysis of the paired subsets from three healthy donors revealed a behavior similar to that already observed in DNA methylation. That is, cMO and ncMO exhibited the most distinct profiles, while iMO displayed an intermediate phenotype. In particular, 3035 genes presented differential expression between at least two of the subsets (differentially expressed genes, DEG, adjusted P-value < 0.05 and Fold change (FC) < 0.5 for downregulated or > 2 for upregulated) (Figure 7A, Supplementary Figure 2A). Specifically, 1617 genes had decreased expression levels in the differentiation from cMO to ncMO (cluster E1), while 1412 DEGs showed increased expression levels (cluster E2).

GO analysis of the DEGs showed that cluster E1 genes (overexpressed in cMO) were enriched for pathways related to functions previously described for cMO subset.

Response to bacterium, response to lipopolysaccharide or angiogenesis were among the GO terms enriched in this cluster. Cluster E2 genes, on the other hand, were enriched in functions repeatedly related to T cell activation and cell-cell adhesion which is also concordant with previously described functions for the ncMO subset (reviewed in [29] and [48]). Interestingly, in agreement with the analysis of ontology terms of the DNA methylation data, interferon gamma-production related terms appeared significantly enriched in cluster E2 (Figure 7B).

Using DoRothEA's datasets (Discriminant Regulon Expression Analysis [203]) we predicted TFs activity through the expression of their target genes (Figure 7C) in the differentiation from cMO to ncMO. This analysis identified Hypoxia Inducible factor 1 (formed by a heterodimer of HIF-1α and ARNT) and PU.1 (*SPI-1*) as the TFs with the highest activity in cMO. PU.1 is a key regulator of monocyte differentiation, also predicted by motif enrichment analysis of CpGs hypomethylated in cMO (Figure 7G, cluster M2). HIF1 is a crucial factor that regulates cellular response to low oxygen concentrations or hypoxia. Also in monocytes, HIF-1α regulates many aspects of the cell's phenotype and metabolism under these stress conditions [218]. The observed decrease in HIF-1 regulons, as well as the expression of the factors themselves (Figure 7D and Supplementary Figure 2B), could potentially be explained by the increase in oxygen concentrations encountered by monocytes during their differentiation process. In the bone marrow, where cMO originate, there is less than 5% of oxygen concentration. In the bloodstream cMO only spend 24h, during which the oxygen concentration varies between 12-20%. Conversely, ncMOs have resided in the bloodstream for approximately 10 days, during which they never encounter hypoxic conditions [219,220]. Therefore, it is expected that cMOs have a phenotype adapted to hypoxic conditions, regulated by HIF1, which is lost during the differentiation towards ncMOs. Other TFs that appear to have an enriched activity in cMO based on the differential expression between cMO and ncMO include NF-κB, AhR and MAFB. These TFs are very important for the activation of monocytes under different conditions.

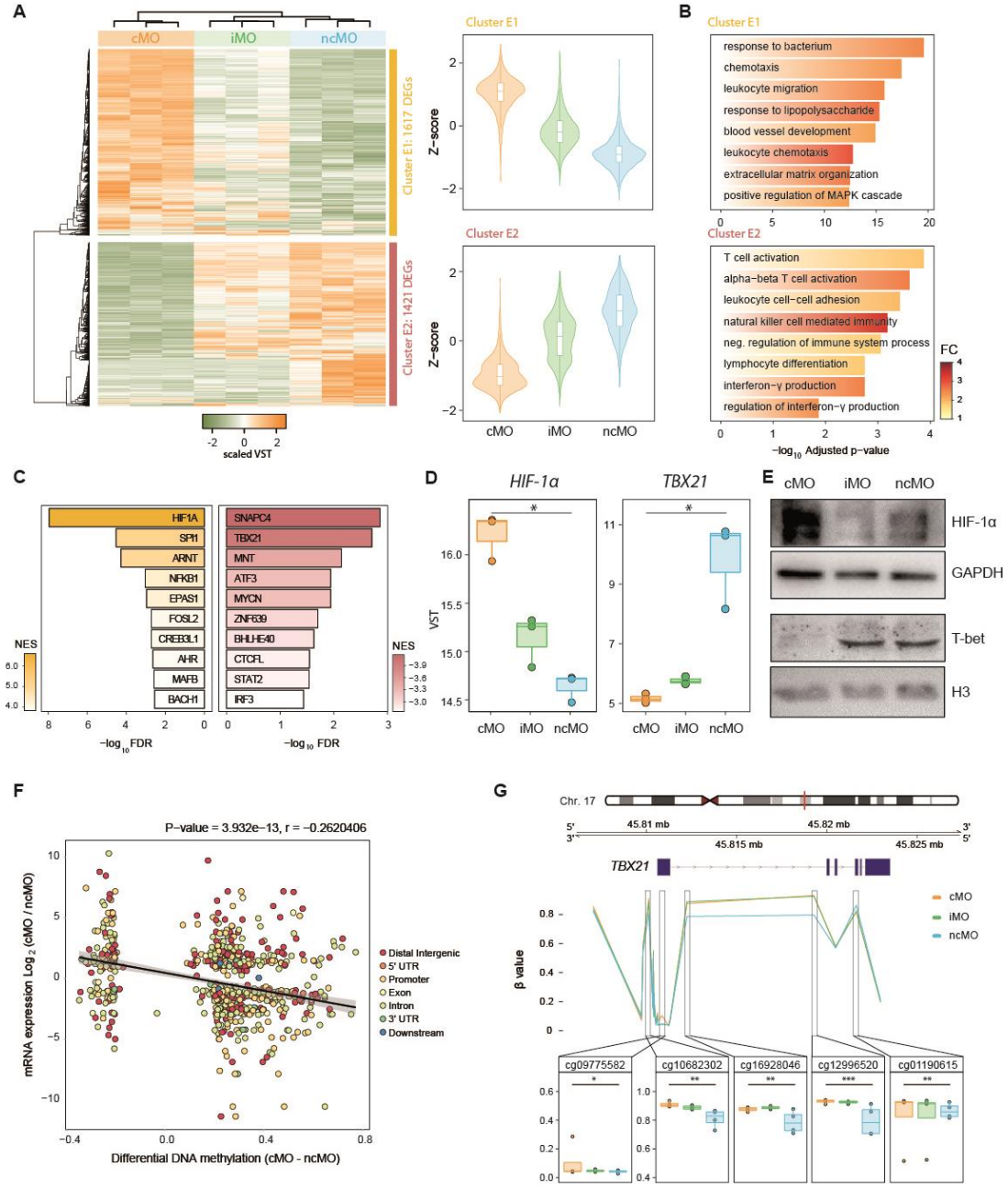


Figure 7. Transcriptomic profile of monocyte subsets at steady state. A. Heatmap plot and cluster analysis (left) of differentially expressed genes (DEGs) between at least two monocyte subsets (FDR < 0.05, fold change < 0.5 or > 2). The color annotation of the lateral bar represents the association with cluster E1 in orange (downregulated in ncMO) or cluster E2 in red (upregulated in ncMO). Violin plots (right) of the DEGs normalized values separated by cluster association. **B**. Gene ontology terms associated with the DEGs in each cluster. Bar length represents log-transformed binomial q values of the term enrichment. Bar color correlates with term's fold change (FC). **C**. Bar plot of the TFs with the highest predicted

activity based on their target genes in each subset in the comparison cMO-ncMO as predicted by DoRoThEA. Color represents normalized enrichment score (NES) of each TF. **D**. Boxplot representation of the expression of *HIF-1α* and *TBX21* genes in each monocyte subset. **E**. Representative western blot of the expression of HIF-1α and T-bet in the sorted monocyte subsets. **F**. Dot plot of the correlation between the differences in methylation in the DMPs and the differences in their associated genes in DEGs. Color represents the position of the CpG in the gene context. **G**. Graphical representation of the methylation status of the CpGs surrounding *TBX21* gene. Statistical significance: * is p-value < 0.05. Two-tailed Wilcoxon's test (D).

Interestingly, regulon analysis of the ncMO transcriptome revealed the participation of SNW1-associated protein 4 (*SNAPC4*), a factor relevant for mRNA splicing, T-box TF *TBX21*, a well-known lineage defining TF from naïve Th precursors to Th1 [221] and a Max-binding protein from the network Myc/Max/Mad called *MNT*. Considering these results, reports associating ncMOs with T cell regulation and communication and given that *TBX21* is the only one among the top predicted TFs to exhibit differential expression between cMOs and ncMOs (see Figure 7D and Supplementary Figure 2B), we decided to further investigate the role of *TBX21* in this differentiation process.

TBX21 encodes the TF known as T-bet. This TF is preferentially expressed in T cells and NK cells. In fact, previous studies showed that it is not expressed in monocytes at basal state, but its expression can be induced upon treatment with IFN-γ [222]. Other studies indicate that it is overexpressed in human M1 macrophages [223]. *TBX21* was shown to be significantly overexpressed in ncMO in comparison to cMO in our transcriptomic data (Figure 7D). These results were validated at the protein level, indicating a potential contribution of T-bet in ncMO, and maybe in iMO, that would be differential to cMO (Figure 7E). The antagonistic activity of HIF1 and T-bet in the subsets was also validated by the differential expression of most of their known target genes (Supplementary Figure 2C), suggesting a contribution of these factors to the differential transcriptome of the monocyte subsets.

We then performed an integration of our DNA methylation and expression data to test whether DNA methylation can regulate gene expression in our datasets. A

significant inverse correlation was seen between DMPs-associated genes by proximity and DEGs (P-value = 3.92e-13, correlation coefficient [r] = - 0.262) (Figure 7F). Genomic context of the DMPs has been proved to be an important factor to consider when studying its impact in gene expression regulation [224]. For this reason, we studied correlation in each of the tiles we use to divide genomic annotations. Only promoter and intronic associated DMPs showed significant correlation with their annotated gene's expression (promoters: P-value = 7.576e-9, r = - 0.3961, see Supplementary Figure 2D for example genes; introns: P-value = 2.771e-8, r = - 0.3164). Considering this, we studied the methylation status of the CpGs proximal to *HIF1-α* and *TBX21* genes. Many of them showed significant differences in methylation (p-value < 0.05) indicating that the differential expression of these factors might be regulated at the epigenetic level (Figure 7G and Supplementary Figure 2E).

4.1.3 T-bet as an important TF for the differentiation from cMO to ncMO

Based on our results, which highlight the significance of the TF T-bet in the transcriptomic profile of ncMO (Figure C, right panel) and the ontology terms relating this transcriptomic profile to that of T cell activation (Figure 7B, cluster E2), we hypothesized that T-bet might have a role in the differentiation from cMO to ncMO. T-bet can be induced by a variety of extracellular signals. In NK cells, it has been proven to be induced in response to cytokines such as IFN-γ, interleukin-12 (IL-12), IL-15 or IL-21 (reviewed by Huang et al in [225]). To test whether monocytes were equally responsive to these stimuli in terms of induction of T-bet, we performed stimulations of monocytes *in vitro* for 6 and 24 hours (Figure 8A). Our results showed that T-bet in monocytes is mainly induced in response to IFN-γ, both at the mRNA (Figure 8B) and protein (Figure 8C) levels, which is in agreement with previous results [222]. The mRNA increase is higher at 6 than 24 hours while the protein increase is more pronounced at 24 hours. Moreover, this induction probably takes place through the IFN-γ canonical pathway, which involves STAT1 phosphorylation (Figure 8C).

Taking advantage of available high throughput public data of the response of monocytes to IFN-γ, we studied whether part of the differentiation process from cMO to ncMO could be explained by this stimulus. In terms of DNA methylation, we used the data from GSE134425, where they stimulated PBMCs with several stimuli for the posterior analysis of the methylome of the sorted stimulated monocytes. In this study, they tested three stimuli: IFN-γ, IFN-α and TNF-α. By comparing their results to our data, we observed that ncMO are significantly more similar to the stimulated samples in all instances. In this regard, the positions that become demethylated in either of the three stimuli also display demethylation in ncMO. This is particularly relevant for IFN-γ (Figure 8D, E). Unexpectedly, this was not the case for expression data. For this analysis, we made use of the data from GSE130567 where they compared the effect of IFN-γ to primary monocyte-derived macrophages. In this case, we found that the genes significantly upregulated during this stimulation were significantly more expressed in cMO than in ncMO (Figure 8F, G). From these analyses, we conclude that ncMO have an epigenetic profile similar to stimulated monocytes. However, transcriptomically their phenotype is not as straightforwardly explained as an individual stimulus response.

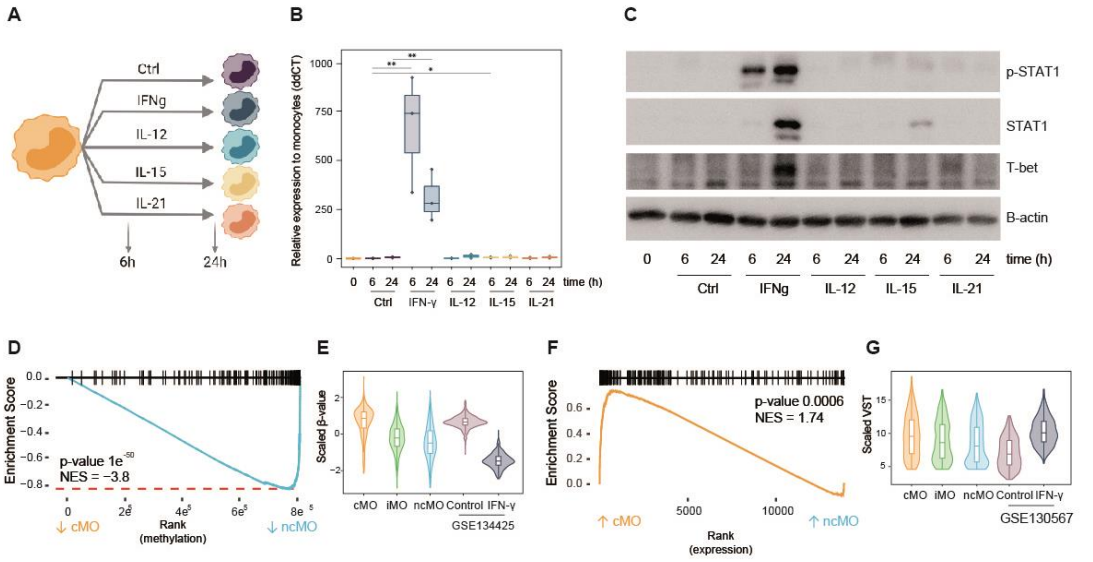


Figure 8. Induction of T-bet in monocytes. A. Graphical representation of the different stimuli used for the induction of T-bet expression in monocytes. **B**. Barplot representation of

TBX21 mRNA expression levels upon stimulation with the various molecules. Data is normalized to housekeeping gene RPL38. **C**. Representative western blot of induction of T-bet and STAT1 protein levels upon stimulation with various molecules. **D**. Gene Set Enrichment Analysis (GSEA) of CpGs with demethylation associated to IFN-γ stimulation of monocytes in the dataset GSE134425. Analysis performed within the differentiation from cMO to ncMO. **E**. Violin plot representation of the methylation status of the CpGs used for GSEA analysis in D. **F**. GSEA analysis of the genes with upregulation in the treatment of monocyte-derived macrophages (M1) with IFN-γ in the dataset GSE130567. Analysis performed within the differentiation from cMO to ncMO. **G**. Violin plot representation of the expression of the genes used for GSEA analysis in F.

4.1.4 T-bet deficiency affects the whole monocyte compartment

To investigate the role of T-bet in the monocyte compartment, we investigated the impact of a homozygous mutation in T-bet in an individual in comparison to a heterozygote male progenitor, used as a control, by inspecting the single-cell RNA-seq datasets from their peripheral blood (Figure 9A) [205]. The affected individual suffers from Mendelian susceptibility to mycobacterial disease, which was directly attributed to the indel variant in homozygosity in the *TBX21* sequence, replacing two amino acids that are highly conserved across species. This causes a loss-of-function allele that, although allows expression of T-bet itself, is unable to induce the expression of its target genes. The first effect we observed at the single-cell RNAseq level was a variance in the proportions of different cell types between the affected individual and his control progenitor (Supplementary Figure 3A). As expected, there is a marked reduction in the contribution of the mutant individual on CD8 Naive T cells and NK cell populations. These are two cell types where T-bet factor is known to be crucial for their development [225–227]. More surprisingly, we discovered an equal reduction in the myeloid compartment, suggesting a role for this factor to the proper development of this lineage as well.

By focusing on the deficient myeloid compartment, we identified clusters of cells with a transcriptomic signature similar to that observed in our bulk RNA-seq analysis. The cMO and ncMO signatures were defined as the top 200 genes with the lowest p-value in the comparison between the two subsets, in each direction of fold change. These signatures identified a region of the object clearly expressing ncMO signature, while the vast majority of the myeloid compartment was more enriched for the cMO

signature (Figure 9B). Interestingly, this region forms a distinct cluster that differentiates from the rest of the myeloid compartment even at very low resolutions, and it is notably less abundant in the affected individual compared to the control progenitor (see Figure 9C). Additionally, there is another cluster of cells in the cMO region that becomes apparent at higher resolution and is also deficient in the mutant individual.

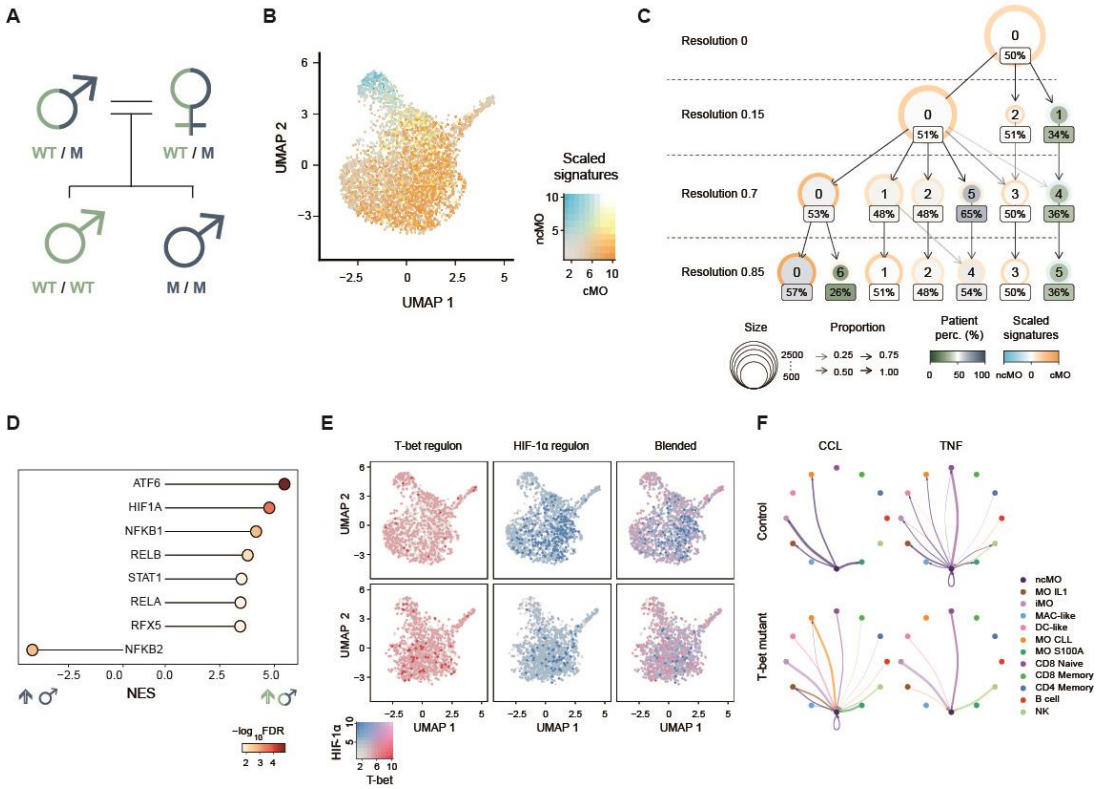


Figure 9. Integration and exploration of single-cell RNA-seq dataset containing a homozygous mutant for *TBX21*. **A**. Schematic representation of the genealogical circumstances of the samples used for single-cell RNA-seq analysis [205]. The samples used were the homozygous mutant progeny and the heterozygous male progenitor. **B**. Uniform Manifold Approximation and Projection (UMAP) representation of the myeloid compartment. Cell color represents the score combination for the cMO and ncMO signatures generated from the top differentially expressed genes in the RNA-seq bulk dataset. **C**. Representation of unsupervised clustering results at different resolutions of the myeloid fraction of PBMCs from the single-cell RNA-seq dataset EGAS00001004504. The node internal color represents the percentage of T-bet mutant's contribution to each cluster. The node external color represents the cMO and ncMO signature score of the cluster. The transparency of the edge shows the incoming node proportion. **D**. Bar representation of the TFs with the highest predicted activity of control vs mutant cells based on their target genes as predicted by DoRoThEA. Bar length represents the normalized enrichment score (NES) of each TF. Circle color represents the

logarithm of the adjusted p-value. **E**. UMAP representation of HIF-1α and T-bet target genes expression in the myeloid compartment of the control (top row) and affected individual (bottom row). **F**. Circle graph representation of CellChat's estimated cell-cell communication between the ncMO cluster of cells and the rest of cells in the single-cell object.

We then performed a discriminant regulon expression analysis between the control and the affected individual in the whole myeloid compartment with the DoRothEA dataset [203] (Figure 9D). This analysis revealed that one of the top TF responsible for the differences was HIF-1α. This suggest that the deficiency of T-bet activity also affects the proper activation of the HIF-1 pathway so that in the absence of T-bet there is also a reduction of the expression of HIF-1α target genes (Figure 9E and Supplementary Figure 3B). Other TFs that appeared to be negatively altered in the absence of T-bet involved the NF-kβ pathway with components such as NFKB1, RELB and RELA. This pathway is very tightly associated with HIF signaling in macrophages [228].

Finally, we studied what impact could T-bet dysfunction have on the phenotype of these monocytes. To this end, we studied the predicted interactions they would have with the rest of the immune cells. We focused our study on the effect T-bet deficiency has on the ncMO-like cluster of cells, due to our demonstration in bulk of T-bet expression in cells with this transcriptomic profile. For this analysis, the myeloid object was divided into several clusters that were annotated based on their top differential markers and the functions enriched within them (Supplementary Figure 3C-E). By using CellChat software [209], which explores cell-cell communication based on the expression of ligand-receptor pairs, we identified a few cell communication pathways that were altered in the individual carrying the homozygous mutations. In particular, the affected individual's ncMO cells expressed higher levels of CCL signaling pathways receptors, making them more susceptible to receiving this type of signaling from a broad plethora of immune cell types. This includes a potential positive feedback loop among themselves. On the other hand, these affected ncMO appeared to have fewer levels of TNF signaling than their control counterparts, making them less able to respond to this inflammatory signaling (Figure 9F and

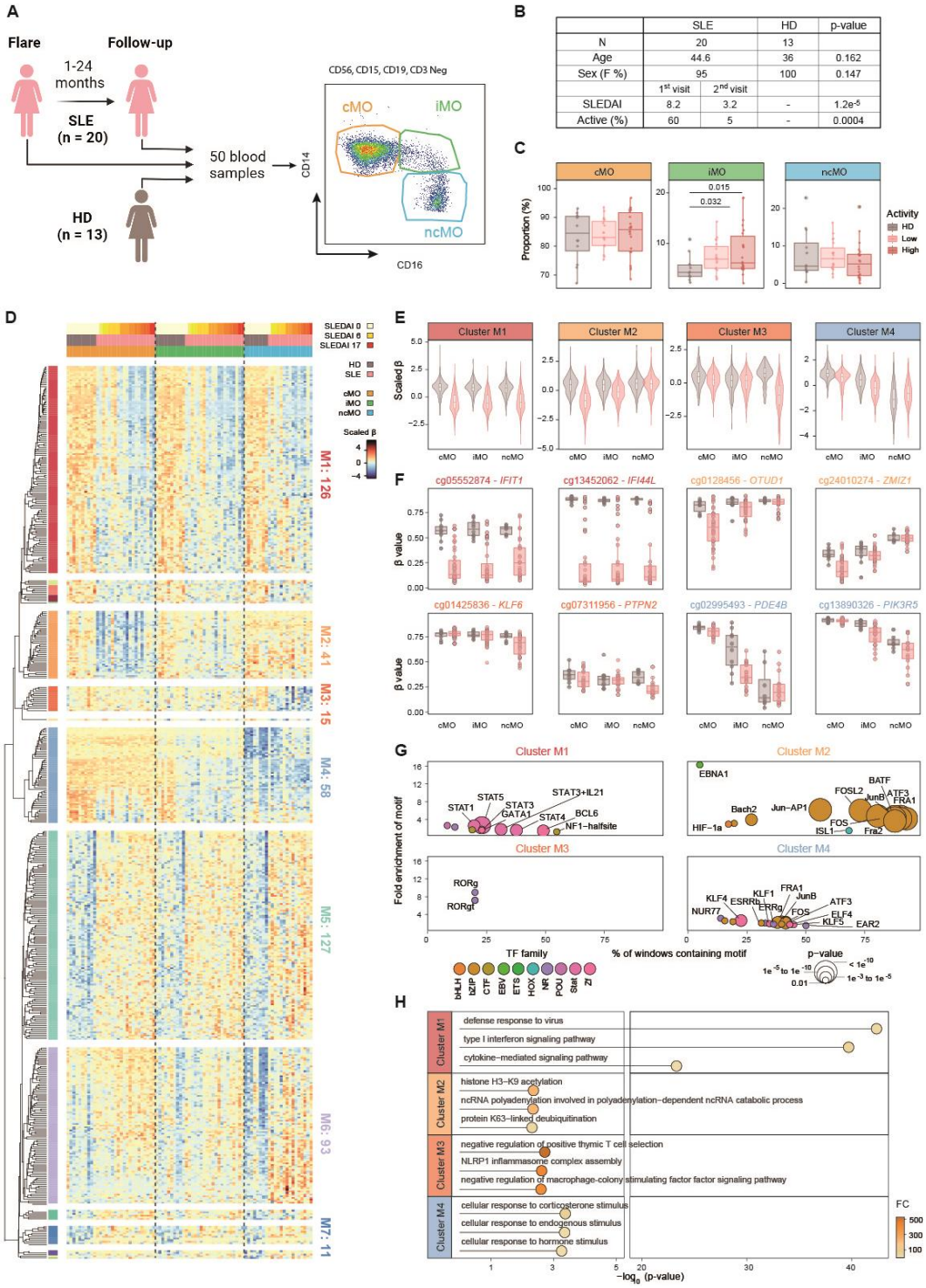
Supplementary Figure 3F). Interestingly, one of the pathways responding to TNF signaling is the canonical NF-kβ pathway [229,230] which would correlate with the previous analysis where multiple components of this signaling pathway were altered in the comparison between the control and the affected individual.

4.2 PART 2. Epigenomic and transcriptomic changes in monocyte subsets in SLE

4.2.1 Monocyte subset-specific DNA methylation profiles in SLE reveal a divergent shift in their epigenetic programs

We collected peripheral blood samples from 20 patients with SLE during their initial hospital visits due to a flare episode of the disease or an increase in their symptoms. Subsequently, we collected new blood samples from these same individuals during their follow-up visits to the doctor. These follow-up visits occurred within a period ranging from 1 to 24 months after the first visit, with 15 patients achieving remission. Additionally, we collected blood samples from 13 age-matched and sex-matched HDs (Figure 10A,B). Blood samples were processed for the isolation of the three monocyte subsets through a flow cytometry cell sorting strategy similar to previously published studies [90,174,175]. In brief, after obtaining the monocytic fraction by negative selection, monocyte subsets were separated based on the surface expression of CD14 and CD16 (Figure 10A).

The percentages of the obtained monocyte subsets were analyzed, revealing a significant increase in the proportion of iMOs in patients with SLE. Notably, this difference appeared to be more pronounced in patients with low disease activity (SLEDAI 6). The proportion of ncMOs also exhibited differential trends between these two groups of patients, once again showing relatively higher levels in patients with low disease activity (Figure 10C). This finding was consistent with previous reports, which have indicated an increase in CD16+ monocytes. The association of these results with disease activity could help explain the disparity observed in previous reports, where some describe an increase in iMOs [88,89] while others reported an increase of ncMOs [90,91].



	SLE		HD	p-value
N	20		13	
Age	44.6		36	0.162
Sex (F %)	95		100	0.147
	1st visit	2nd visit		
SLEDAI	8.2	3.2	-	$1.2e^{-5}$
Active (%)	60	5	-	0.0004

Figure 10. Epigenetic profiles of monocyte subsets in SLE. A. Schematic representation of the cohort studied and the acquisition of the monocyte subsets. **B**. Summary table of the

cohort characteristics. **C**. Boxplot representation of the percentage of monocyte subsets within the monocyte fraction of PBMCs. Contains all samples collected, including second visits of SLE samples. Low activity are samples with SLEDAI of 5 or lower, high activity are samples with SLEDAI of 6 or higher. Statistics performed with two-tailed Wilcoxon test. **D**. Heatmap representation of the 499 differentially methylated positions (DMPs) in either of the pairwise comparisons between SLE and HD samples from each subset, only first visit included. **E**. Summary violin plot representation of the first four clusters of DMPs. In grey, HD samples; in pink, SLE. **F**. Illustrative examples of two individual DMPs from the first four clusters. In grey, HD samples; in pink, SLE. The color of the CpG name represents the cluster it belongs to. **G**. Motif enrichment analysis of the sequence englobing the DMPs. **H**. Gene Ontology enrichment results of the genes annotated closest to the DMPs from the first four clusters.

We subsequently generated the DNA methylation profiles of cMOs, iMOs and ncMOs from both the SLE patient and HD cohorts. Our analysis revealed unique differences in DNA methylation patterns among the different monocyte subsets when comparing patients with SLE with HDs. Only the first visits of SLE samples were used in the upcoming analyses to avoid potential biases from biological replicates. Specifically, with an absolute differential of beta value >0.1and an FDR <0.2, cMOs displayed 289 DMPs, iMOs exhibited 118 DMPs and ncMOs had 201 DMPs (Figure 10D). These produced a total of 499 unique DMPs, with some overlap between monocyte subsets. These DMPs were then grouped into seven clusters (M1–M7), based on the DNA methylation levels of the different samples. We identified clusters of DMPs exhibiting a similar behavior across all monocyte subsets and clusters of DMPs displaying a distinctive profile in one of the subpopulations (Figure 10E and Supplemental Figure 4A). For instance, cluster M1 consists of 126 DMPs that are hypomethylated in SLE compared with HD across all three monocyte subpopulations. Examples of these DMPs include CpGs annotated to IFN-related genes, such as *IFIT1* and *IFI44L* (Figure 10F). Cluster M2 (41 DMPs) annotate to genes that have a more drastic hypomethylation in cMOs than in both iMOs and ncMOs when comparing HDs versus SLE (Figure 10D,E). It is the case of genes like *ZMIZ1*, a member of the protein inhibitor of activated STAT protein family, and *OTUD1*, a deubiquitinase related to TNF and IFN signaling [231] (Figure 10F). Remarkably, cluster M3 (15 DMPs) corresponds to positions hypomethylated in SLE ncMOs in comparison with HD ncMOs (Figure 10D,E). Examples include those annotating at

genes like Krupple-like family transcription factor (TF) *KLF6*, which has been related to the aryl hydrocarbon receptor pathway, and *PTPN2* (Figure 10F). Cluster M4 contains 58 DMPs that demethylate in the differentiation from cMOs to ncMOs and that in SLE iMOs are more advanced in the demethylation process than their HD counterparts (Figure 10D, E). Examples of DMPs in the M4 cluster include those annotating to *PDE4B*, which encodes a key element for the monocyte's response to LPS [232] and *PIK3R5*, which encodes a regulatory subunit of PI3K complex, relevant for several immune functions in monocytes including cytokine release and adhesion [233–235] (Figure 10F). The remaining clusters from M5 to M7 (127, 93 and 11 DMPs) were characterized by hypermethylation in SLE in comparison with HDs, being cluster M5 the predominant, with similar hypermethylation levels for all monocyte subsets (Supplemental Figure 4A,B). The analyses of differentially methylated regions (DMRs) produced overlapping results with 60, 27 and 41 DMRs in cMOs, iMOs and ncMOs, respectively (data not shown). In summary, these results suggest that dysregulation of DNA methylation in SLE affects monocyte subsets in different ways. Since iMOs derive from cMOs and ncMOs from iMOs, one can interpret that some determinants in SLE pathology are affecting the differentiation process at the DNA methylation level.

Annotation of the DMPs to the genome in relation to CpG islands showed different patterns in various clusters (Supplemental Figure 4C). For instance, clusters M1 and M3 showed an enrichment in CpGs annotated to shore regions in comparison with the background. In contrast, the remaining clusters presented a higher representation of positions outside CpG islands. In parallel, annotation of the DMPs in relation to the gene location also separated cluster M1 from the rest (Supplemental Figure 4D). This showed a significant and marked representation of promoter regions encompassing more than 60% of the DMPs in this cluster. Cluster M3 also showed a high representation of promoter regions as well as exons. The remaining clusters showed a more homogeneous representation with similar region percentages to those present in the background. This highlights the common behavior of the subsets

in demethylating promoters, while subset-specific changes appear to participate in more complex regulatory processes.

TF-binding motif enrichment analysis in the regions surrounding DMPs showed that the different clusters exhibit significant enrichment for a wide range of TFs (Figure 10G, Supplemental Figure 4E). For instance, DMPs from cluster M1 are enriched for binding motifs of several STAT TFs, indicating their relation to immune cell activation pathways. Additionally, approximately 80% of DMPs from cluster M2 (hypomethylated in cMOs from patients with SLE) presented a significant enrichment for TFs of the Fos and Jun family. These have been associated with monocyte-to-macrophage differentiation and activation [236], as well as with other inflammatory diseases such as rheumatoid arthritis (reviewed in Zenz et al [237]). We were particularly interested in DMPs from cluster M3 because they are hypomethylated in SLE ncMOs, less studied than cMOs. DMPs from cluster M3 presented an enrichment for the consensus sequence of RORγ, which is a TF involved in Th17 differentiation [238]. This factor has also been described in a subset of monocytes which is associated with IL-17 production in pathological conditions [239].

We also performed GO analyses of the genes associated with these DMPs. Hypermethylated DMPs in SLE in comparison with HDs in either cMOs or ncMOs were associated with metabolic synthesis and degradation of several compounds such as ‘spermidine biosynthesis process’ or ‘allantoin metabolic process’ (Supplemental Figure 4F). Notably, the polyamine spermidine has been previously described to be present at decreased concentrations in the plasma of patients with SLE [240]. Also, allantoin is the product of uric acid non-enzymatic oxidation. Urate may be found at elevated levels in the serum of patients with active SLE where it facilitates the activation of inflammatory pathways, particularly in those with kidney damage [241]. On the other hand, DMPs that lose methylation in SLE compared with HDs in either cMOs or ncMOs were annotated to genes related to immune system activation. DMPs exhibiting hypomethylation in cMOs were strongly associated with ‘type I IFN signaling pathways’, a proinflammatory group of cytokines that are a key player in

SLE development and pathogenesis. This is consistent with previous results describing significant hypomethylation in IFN pathways in SLE immune cells [77,150,161,242,243].

We also conducted GO enrichment analysis of these positions when clustered in an unsupervised manner as depicted in Figure 10D (see Figure 10H, Supplemental Figure 4G). The results showed that DMPs in cluster M1, which are hypomethylated in SLE in all three monocyte subpopulations, strongly associated with type I IFN response pathways consistent with our previous analysis. Notably, cluster M3, which exhibited significant hypomethylation in SLE within the ncMO subset, was associated with pathways typically associated with T cells and pathways related negatively to monocyte differentiation to macrophages. This highlights once more the distinct involvement of ncMOs in SLE.

4.2.2 Monocyte subset-specific transcriptomic alterations in SLE

We then performed RNA-seq analysis of 60 samples, comprising the three monocyte subsets of 13 patients with SLE during their first visit and 7 HDs. The analysis revealed 2805 DEGs in cMOs, 1916 in iMOs and 1287 in ncMOs s (FDR < 0.05 and $\log_2$FC > 1 or <−1). In general, the majority of DEGs corresponded to genes with higher expression levels in SLE compared with HDs (Figure 11A, B). Also, most changes presented a similar behavior in the three monocyte subsets, although a high percentage of the total DEGs did not reach statistical significance in ncMOs (Figure 11B). DEGs were divided into seven main clusters (E1–E7) in an unsupervised manner based on their expression levels (Figure 11B, C and Supplemental Figure 5A). DEGs in E1 (986 DEGs) and E2 (631 DEGs) corresponded to genes that showed increased expression in SLE versus HDs in all subsets, with a more pronounced effect in cMOs. Notably, genes in cluster E2 displayed a progressive decrease in expression during the differentiation from cMOs to ncMOs. In cluster E3 (151 DEGs), genes showed similar expression levels in the three monocyte subsets in HDs and a similar upregulation in the three subsets in patients with SLE. In cluster E4 (70 DEGs), genes were downregulated during differentiation from cMOs to

ncMOs but had a particular increase in expression in ncMOs of SLE. Cluster E5 (971 DEGs) and E6 (375 DEGs) showed a loss of expression in the three subsets in SLE. Finally, cluster E7 (322 DEGs) showed a parallel increase in expression during the differentiation from cMOs to ncMOs in both HDs and SLE.

Some genes with a unique behavior in one of the subsets are represented in Supplemental Figure 5B. Among these, *CD58* and *IL18R1* displayed a marked upregulation in cMOs from SLE. CD58 is a molecule expressed in the cell surface of monocytes, playing a crucial role in immune synapsis formation with CD2-expressing T and natural killer cells [244]. Dysregulation of this axis, characterized by aberrant expression of CD58, has been previously linked with several autoimmune diseases including rheumatoid arthritis [245,246]. On the other hand, the proinflammatory cytokine IL-18 has increased levels in the serum of patients with SLE [247] and the expression of IL18R in myeloid cells has been linked with their ability to migrate and be recruited to the site of inflammation [248]. The TF IRF4, which we found specifically upregulated in SLE iMOs, is expressed in myeloid cells upon stimulation with IFN-β and it induces their activation and differentiation [249]. *ITGA9*, associated with infiltration and migration in macrophages [250], was found to be especially upregulated in iMOs from patients with SLE, indicating a potential migratory phenotype. In the case of ncMOs, *IFIH1* and *TLR10* were found to have specific differential expression in this subset, indicating a proinflammatory phenotype. *IFIH1* expression leads to a strong inflammatory response involving IFNs, a pathway known to be dysregulated in SLE [251]. In the case of *TLR10*, its expression in monocytes is linked to a suppression of their activation capacity and their ability to activate T cells [252]. In this case, ncMOs from HD increased their *TLR10* expression in comparison with cMOs, but this increase was not present in the patients with SLE which makes them more prone to this interaction than in physiological conditions.

GO analyses revealed enrichment of biological functions similar to those obtained for DNA methylation (Figure 11D and Supplemental Figure 5C). Clusters of DEGs that gain expression (clusters E1–E4) are primarily associated with immune

response functions. For clusters E1 and E2, particularly upregulated in cMOs, functions related to ‘leucocyte migration’ and ‘response to chemokine’ are prevalent.

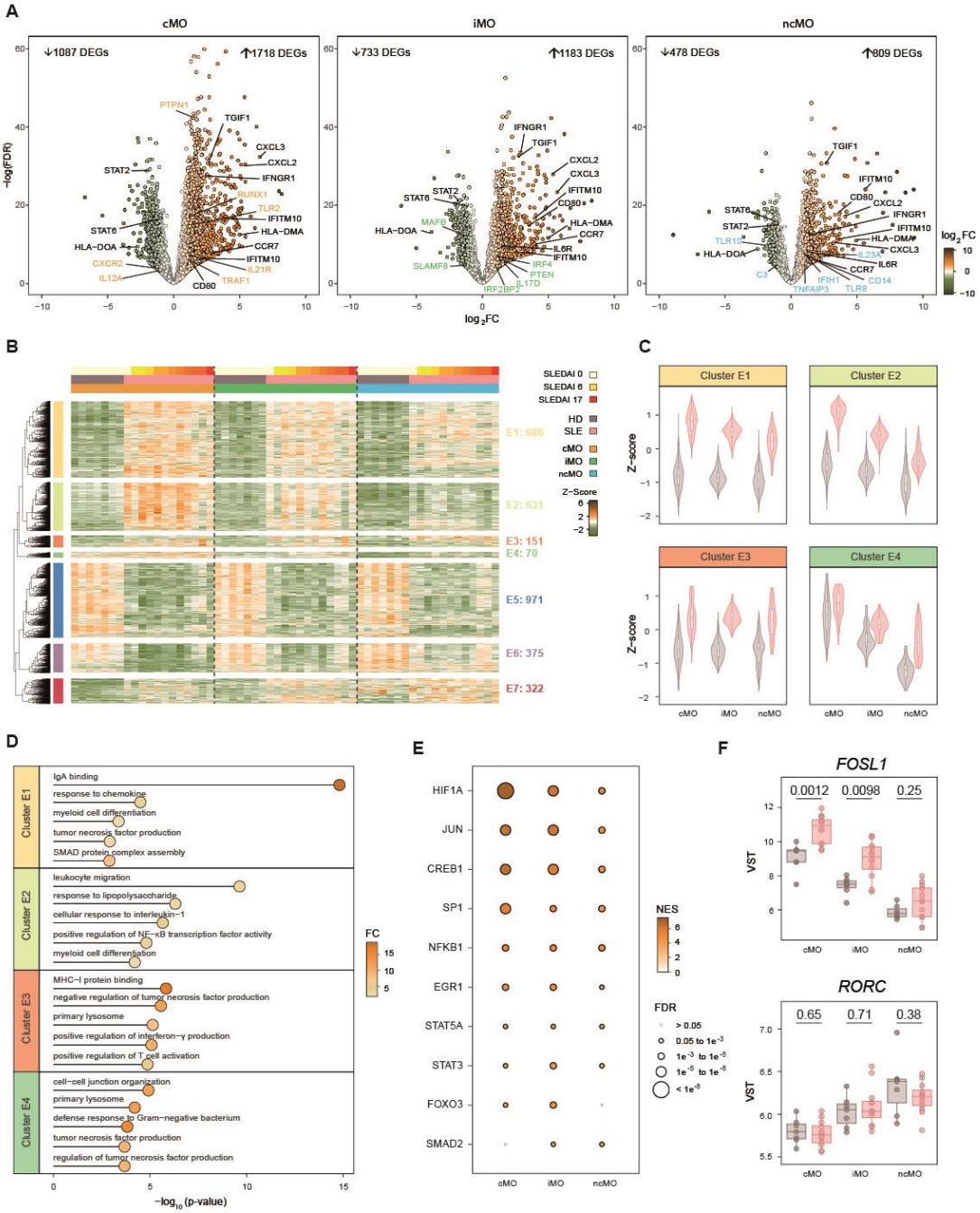


Figure 11. Transcriptional patterns of monocyte subsets in SLE. A. Volcano plot of the differentially expressed genes (DEGs) in each pairwise comparison between SLE and HD

from each subset of monocytes. Labels in black annotate highlighted DEGs common in the three comparisons. Labels in color annotate highlighted DEGs unique to the respective subset. **B**. Heatmap representation of the DEGs from the pairwise comparisons between SLE and HD in either of the subsets. **C**. Summary violin plot representation of the first four clusters of DEGs. In grey, HD samples; in pink, SLE. **D**. Gene Ontology enrichment of the first four clusters of DEGs. **E**. Dotplot representation results of Virtual Inference of Protein-activity by Enriched Regulon analysis. It displays the top 10 transcription factors predicted to have more activity in SLE samples compared with HD samples, individually in each subset. **F**. Expression of highlighted transcription factors resulting from the motif enrichment analysis of DMPs. In grey, HD samples; in pink, SLE. Statistics performed with two-tailed Wilcoxon test.

Also, pathways of response through TNF, IL-1 and NF-κB were highly represented among the genes from these clusters. IFN production and MHC-I signaling are predominant among the functions derived from genes from cluster E3. Cluster E4, similarly to cluster E2, showed functions related to TNF production and response to bacteria molecules. For the clusters of genes undergoing downregulation, we observed enrichment in functions related to autophagy (cluster E5), probably related to the deficit of clearance of autophagocytic residues in SLE [253]. In cluster E6, genes related to several metabolic pathways and to the production of IL-12, a proinflammatory cytokine produced by myeloid cells downregulated by treatment with corticosteroids [254,255], could lead to a T cell switch towards Th17 population [256].

TF involvement was inferred from the gene expression of their targets using VIPER [202]. The TFs with the highest predicted activity in SLE in each of the subsets are shown in Figure 11E. These results were consistent with the observation that the overall transcriptomic profile in SLE is relatively similar among monocyte subsets. HIF-1α is the most significantly enriched TF in each of the subsets. This TF has been tightly linked with TNF-α signaling and inflammatory autoimmune responses (reviewed in Tang et al [257]). SLE monocytes are in a highly activated state, as also indicated by the enrichment of NF-kB, STATs or EGR1 regulons. NF-kB can be activated through TLR signaling or by uptake of micro particles in SLE becoming a therapeutic target [258–260]. EGR1 has also been shown to be associated with inflammatory responses in monocytes [261].

In parallel, we inspected whether the TFs whose binding motifs were enriched in the different DMP clusters (Figure 10G and Supplemental Figure 4C) were also differentially expressed between the populations. This is the case for several of the factors (Figure 11F and Supplemental Figure 5D). For instance, cluster M2 DMPs were clearly enriched for consensus sequences of the family of TFs Fos-Jun. FOSL1 is one of the key members of this family and is differentially expressed in cMOs between HDs and SLE, but not in ncMOs, showing a similar behavior to the DMPs from this cluster. In the case of cluster M3 of DMPs, where only one TF, RORγ, is significantly enriched in the sequences surrounding the CpG positions, its expression is not differential between HDs and SLE in any of the subpopulations. However, it has a clear upregulation between cMOs and ncMOs (Figure 11F), suggesting a more prominent role in the transcriptome regulation of the latter. Other interesting TFs that present differences among the subpopulations include STAT4, ATF3, NR4A1 (*Nurr7*) and IRF4 (Supplemental Figure 5D).

To further investigate the correlation between the epigenomic and transcriptomic profiles of the samples, we performed a GSEA of the genes annotated by proximity to the DMPs from each individual subpopulation (Supplemental Figure 5E, F). The results revealed that the genes annotated to the DMPs from the cMO subset are significantly more enriched in transcriptomic profiles of cMOs from HD samples, in comparison with the SLE samples. Interestingly, cMO's DMPs were also enriched in the iMOs from HD samples, possibly due to the short lifespan of cMOs before they differentiate to iMOs which may cause cMO's epigenomic profile to influence iMO's transcriptomics. ncMO's DMPs did not reach statistical significance threshold for correlation, but they showed a tendency to be more enriched in the SLE samples.

4.2.3 DNA methylation changes correlate with SLE activity and progression

We and others have previously shown that the DNA methylation profiles of immune cells correlate with activity index in rheumatoid arthritis [262–264] and in SLE [265]. To test this potential relationship in the different monocyte subsets, we performed Spearman's correlation between the SLEDAI of the patients at their first visit and their

DNA methylation profiles of the three monocyte subsets. We found 823 CpGs in cMOs, 683 in iMOs and 952 in ncMOs that correlated significantly with disease activity (ρ cut-off=0.7 and p value cut-off=$1\times10e^{-3}$) (Figure 12A and Supplemental Figure 6A,B). These were vastly different between monocyte subsets. However, when analyzing the GO enrichment of the genes annotated to these positions, the results obtained were similar among the subsets. Particularly, the positions with a negative correlation to SLEDAI, that is, positions that are less methylated at higher activity indexes than at low activity indexes, associated with pathways related to immune response via type I IFN pathway in cMOs and iMOs (Figure 12B). These positions did not correlate with activity in the samples from the second visit, which contrasts with what we observed for other autoimmune conditions [262]. As a result, there were very poor correlations when comparing the difference in methylation with the difference in activity ($R^2 < 0.05$, data not shown).

Given the GO enrichment results associating the correlating positions with IFN pathways, together with the potential role of STATs in the acquisition of the observed epigenetic profiles (Figure 10G), we wondered whether these TFs could be associated with the activity in the samples. To address this question, we generated scores for each member of the STAT family for both DNA methylation and expression. For DNA methylation, we calculated the average of the scaled methylation values of all the CpGs with the corresponding STAT consensus sequence in a region of 500bp surrounding the DMPs from cluster M1. For the expression data, we measured the average of the scaled expression values of the target genes of each STAT family member as defined by CollecTRI regulons [204]. The results revealed a significant positive association between the activity index and the levels of STAT targets (see Figure 12C). This association holds true for all members of the STAT family except STAT6, whose targets do not increase their expression along with SLE activity. We highlight the results for STAT1 given that it is the main responder to IFN signaling, and its expression is dysregulated in SLE [266,267]. Samples with higher activity decrease the methylation of STAT1-responsive DMPs while they have a higher expression of STAT1 gene targets (Figure 12D,E). The correlation of

these STAT1-associated DMPs with disease activity was validated with an external public cohort of CD14+ monocytes from 27 SLE donors and 27 HDs (GSE59250; see Supplemental Figure 6D,E). Although the association between the epigenetic STAT1 profile and the activity occurs for the three subsets, the correlation of the transcriptomic profile is only statistically significant in cMOs and iMOs. This suggests that ncMOs are less STAT1-responsive.

Regarding the disease progression of the patients included in the study, 15 out of the 20 patients improved over the following months and achieved a remission state with a SLEDAI below 6 in the second sample collection point. With this information available, we investigated whether the prognosis of the patients could be predicted based on the methylation status of some CpG sites in samples from the first visit. We identified 494 DMPs in cMOs between patients who would remit and patients who would not. With these, we built a supervised learning k-nearest neighbors' algorithm to predict the progression of the disease. With an internal 10-fold cross-validation approach, we confirmed that 3 CpG positions were enough to build an accurate prediction algorithm (Figure 12F, G).

Following the results observed in the first section of this thesis, we set to examine the behavior of T-bet and HIF-1α in the SLE samples. By performing a TF involvement analysis based on the expression of their target genes, we again observed T-bet among the top TF responsible for the transcriptome of ncMOs in comparison to cMOs, both in HD and SLE samples (Figure 12H, right panel). As for the TF most prominent in the cMO signature, HIF-1α appeared on top for the SLE samples. However, in the HD samples, its signature was not statistically enriched in cMO in comparison to ncMO (Figure 12H, left panel). These results were clearly evidenced when analyzing the expression of the target genes in the different types of samples (Figure 12I). The difference in HIF-1α between monocyte subsets was particularly prominent in the high activity samples.

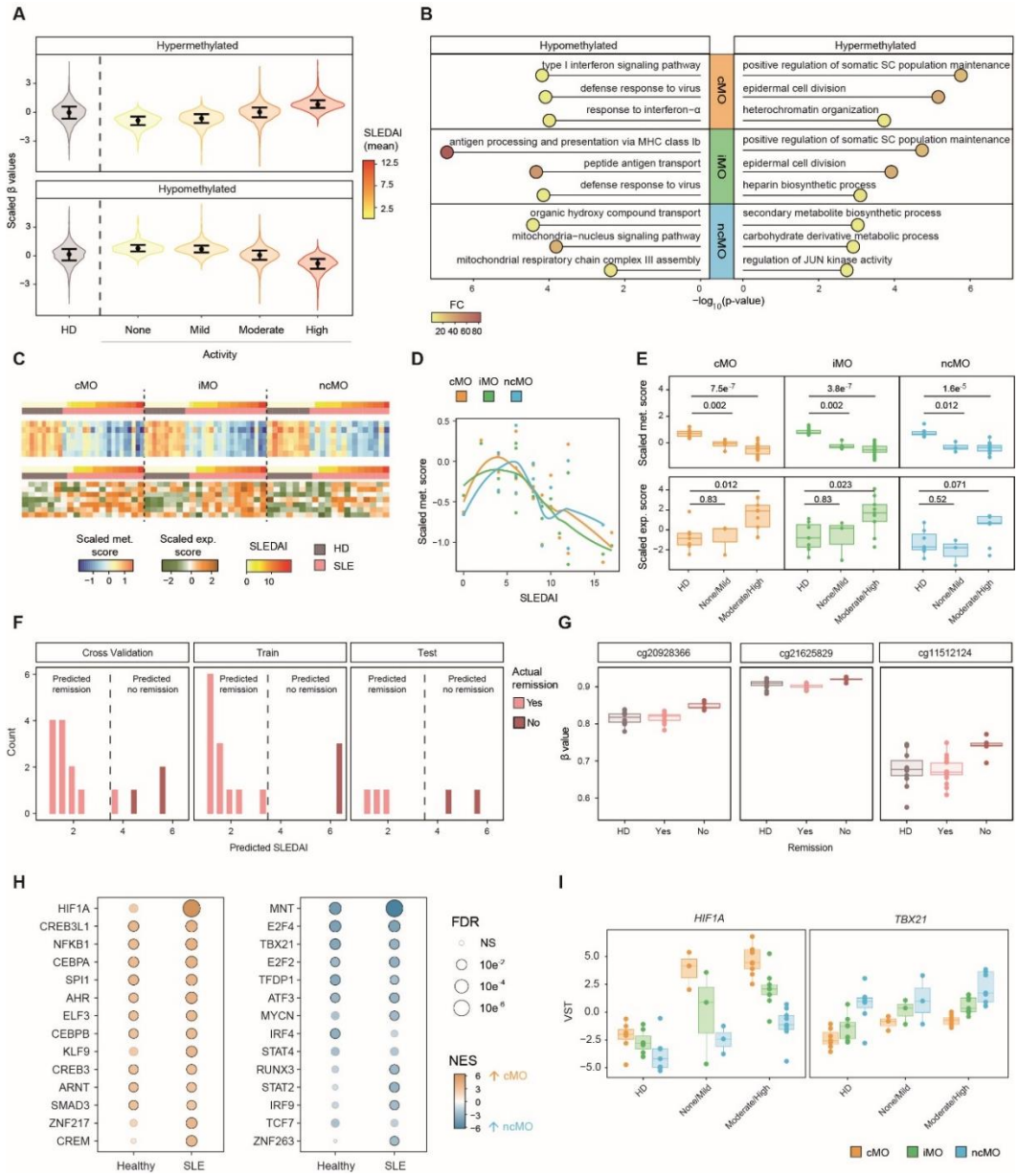


Figure 12. Association of epigenomics and transcriptomics with SLE disease activity. **A**. Violin plot representation of the cMO's CpGs correlating with SLE activity as measured by SLEDAI in a Spearman's correlation. Samples divided in four groups according to their activity index: none is SLEDAI=0, mild is SLEDAI between 1 and 5, moderate is an index between 6 and 10, and high is a SLEDAI above 10. **B**. Gene Ontology results of the genes annotated by proximity to the CpGs correlated with activity in each of the subsets. **C**. Heatmap representation of STAT targets at the methylation (top) and expression (bottom) level. Methylation targets are defined per HOMER database and expression targets as per CollecTRI database. Each row represents the results for a different STAT family member ordered from STAT1 in top row to STAT6 in bottom row. **D**. Dotplot representation of

methylation correlation with SLE activity index of the DMPs that are STAT1 targets in cluster M1. **E**. Boxplot summary representation of STAT1 targets at the methylation (top) and expression (bottom) level in each of the subsets. Methylation targets are defined per HOMER database and expression targets as per CollectRI database. Samples grouped per activity so that none/mild group is SLEDAI 0–5 and moderate/high is SLEDAI above 6. Statistics performed with two-tailed Wilcoxon test. **F**. Barplot representation of the results of the predictive model in the different cohorts. **G**. DNA methylation levels of the three CpGs used by the predictive model in HD, remission patients and non-remission patients. **H**. TF enrichment analysis in the comparison cMO vs ncMO in HD samples and SLE samples. Color represents normalized enrichment score with orange being more enriched in cMO and blue more enriched in ncMO. **I**. Average of HIF-1α and T-bet targets gene expression in each of the subsets. Samples are grouped per activity index so that none/mild group is SLEDAI 0–5 and moderate/high is SLEDAI above 6.

4.2.4 Patients with SLE undergo drastic changes of monocyte subpopulations

To further investigate the presence of differences in our dataset, we leveraged single-cell RNA-seq dataset from PBMC samples of 162 patients with SLE and 98 HDs ([207]-GSE174188). Due to the difficulty of identifying iMOs in single-cell datasets, we initially divided the myeloid fraction of this dataset at a very low resolution, obtaining just two clusters: cluster 0 or CD16− cells and cluster 1 or CD16+ cells (Figure 13B,C). With this initial division, DEGs from our bulk RNA-seq dataset significantly overlapped with the DEGs between SLE and HD cells in the single-cell object (cMO's DEGs in cluster 0, p-value=0.0096; iMO's and ncMO's DEGs in cluster 1, p-value=0.0164) (Supplemental Figure 7A). Some of the genes upregulated in both datasets in the comparison of HD versus SLE included *IRF7, IFI27, HLA-A or HLA-C* (Figure 13D and Supplemental Figure 7B).

The density distribution of cells in the myeloid object exhibited a striking difference between SLE and HD samples. This contrast was especially pronounced in the SLE samples with higher activity (SLEDAI >6) (Figure 13E). This finding prompted us to increase the resolution of the clustering in order to identify clusters containing cells from one specific condition (SLE or HD) (Figure 13F). At a resolution of 0.15 (Figure 13A), two of the new clusters mostly contained cells from the SLE samples. Specifically, clusters 1 and 6 contained 81% and 60% of cells from patients with SLE, respectively (Figure 13A).

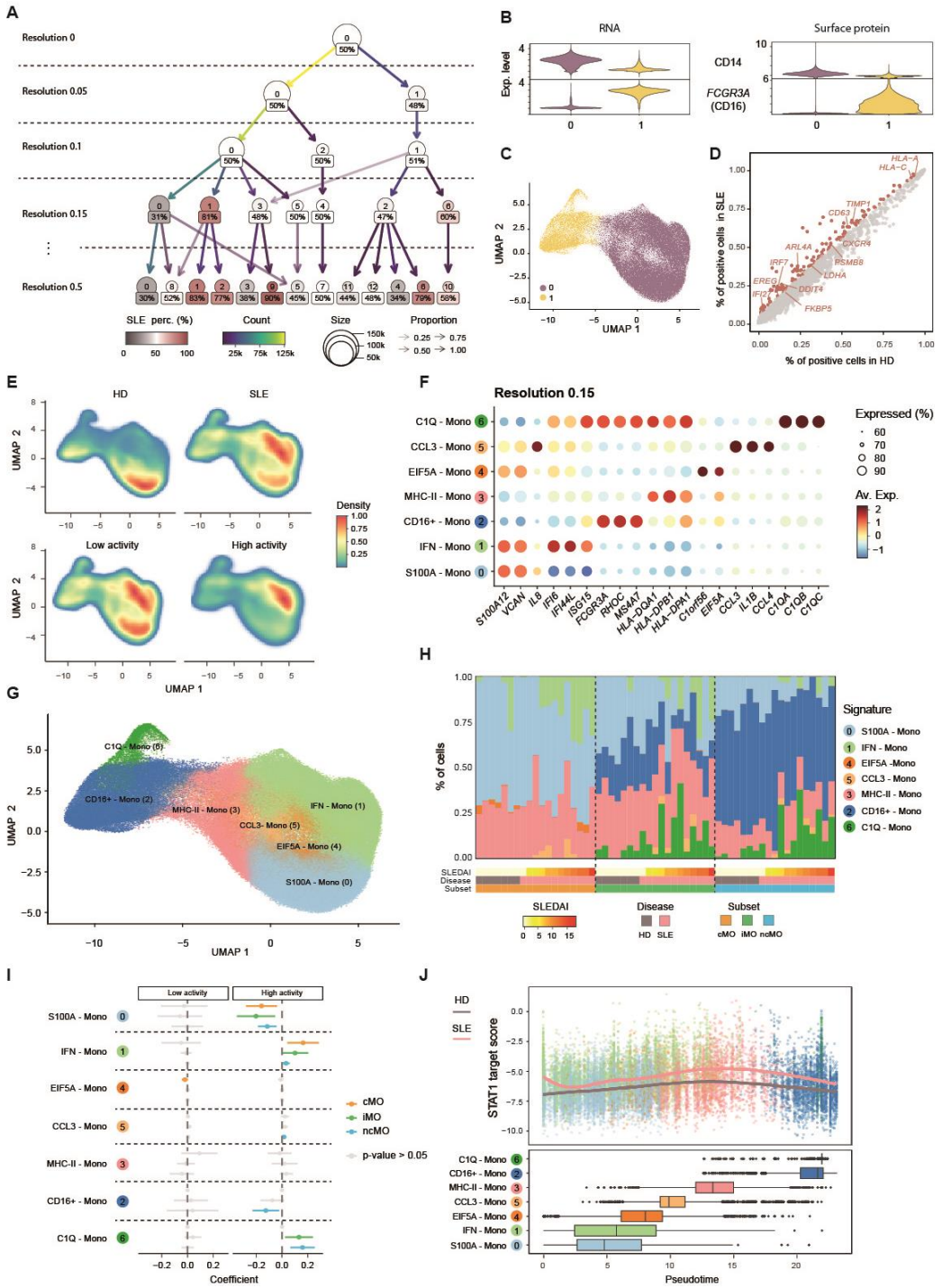


Figure 13. Integration with single-cell RNA-seq object of SLE samples. A. Representation of unsupervised clustering results at different resolutions of the myeloid fraction of PBMCs from the single-cell RNA-seq dataset GSE174188. Node color represents

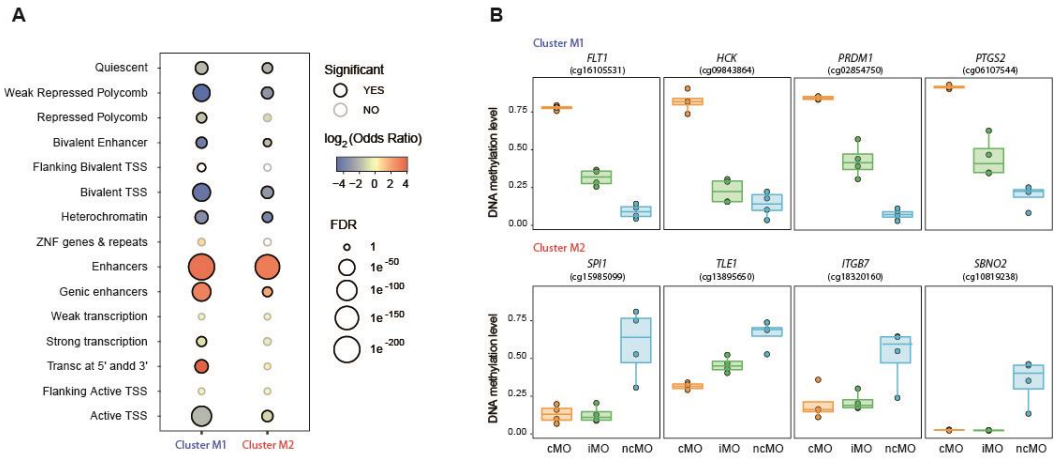
the percentage of SLE composition of each cluster. Edge color represents the number of samples following each division. The transparency shows the incoming node proportion. **B**. Violin plot representation of the expression levels of CD14 and CD16 (*FCGR3A*) at RNA and protein level in the two clusters generated by a 0.05 resolution. **C**. UMAP representation of the distribution of the two clusters generated by a 0.05 resolution. **D**. Scatter plot of the percentage of detected genes in HD and SLE samples in cluster 0 of the resolution 0.05. In color are the DEGs from the single-cell RNA-seq dataset (FDR <0.05, log2FC >0.1). Labels present in DEGs commonly upregulated in the bulk dataset. **E**. UMAP representation showing the density distribution of cells in the HD and SLE cohort (top), and within the SLE cohort, the low activity and high activity samples (bottom). SLEDAI of 6 was used as a threshold. **F**. Dotplot representation of some of the DEGs in each cluster formed by a 0.15 resolution. **G**. UMAP representation of the distribution of the six clusters generated by a 0.15 resolution. **H**. Barplot representation of the estimate percentage composition of each bulk sample according to signature expression from each of the clusters generated by a 0.15 resolution. **I**. Summary plot of percentage composition of bulk samples. Coefficients, p values and 95% CIs derived from fitting a linear model and testing it with Wilcoxon test. **J**. Scatter plot representation of STAT1 targets expression score across the pseudotime distribution of the differentiation from CD16− to CD16+ monocytes (top). Summary representation of pseudotime representation of each cluster (bottom). log2FC, log2 fold change.

Analysis of some of the highly expressed genes for each cluster and their corresponding enriched pathways (Figure 13F, G and Supplemental Figure 7C) led us to the following annotations: cluster 0, primarily formed by cells of HD origin, was labelled as S100A–mono and was enriched in pathways of response to LPS but also of response to oxidative stress. Cluster 1, essentially composed of SLE cells received the name of IFN– mono because it had an unmistakable IFN signature with *IFI6*, *IFI44L* and *ISG15* genes as its top markers. Cluster 2 was the main CD16+ cluster and, therefore, we labelled it as CD16+– mono. Cluster 2 was also enriched in pathways of response to type I IFN and regulation of leucocyte activation. Cluster 3 (MHC-II–mono) was strongly enriched for genes and pathways of the MHC-II and cluster 4 (EIF5A–mono) had higher levels of cell proliferation genes (*EFI4A* and *C1orf56*) and pathways related to cell adhesion and endocytosis. Cluster 5 (CCL3–mono) had a highly proinflammatory signature with marked TNF-α pathway activation. Finally, cluster 6 (C1Q–mono) highly expressed genes and pathways of the complement 1q (*C1QA, C1QB, C1QC*). Interestingly, MHC-II–mono, EIF5A–mono and C1Q–mono presented very similar features to those found in a previously annotated single-cell dataset of another autoimmune disease [268].

By using CIBERSORTx software [213], we studied the proportion of these clusters represented in our bulk dataset (Figure 13H). With this approach, we determined that our bulk samples are composed of a heterogeneous mixture of cells showing differential patterns correlating with disease state and disease activity. As expected, cMOs from SLE donors were composed of a significant percentage of the labelled IFN–mono cluster, while HD samples barely harbored them. This representation was more elevated in samples with higher disease activity, where the difference in proportions was statistically significant. We determined that iMOs contained a mixture of mostly S100A–mono, MHC-II–mono and CD16+–mono in HD samples, while iMO SLE samples repeatedly showed a non-trivial representation of C1Q–mono. This behavior was similar also for ncMO samples, where C1Q–mono has a significantly higher proportion in the SLE samples of our bulk dataset. Another remarkable finding is the presence of the highly proinflammatory cluster CCL3–mono exclusively in SLE samples, being present in all three subsets from the bulk but never in HDs (Figure 13I).

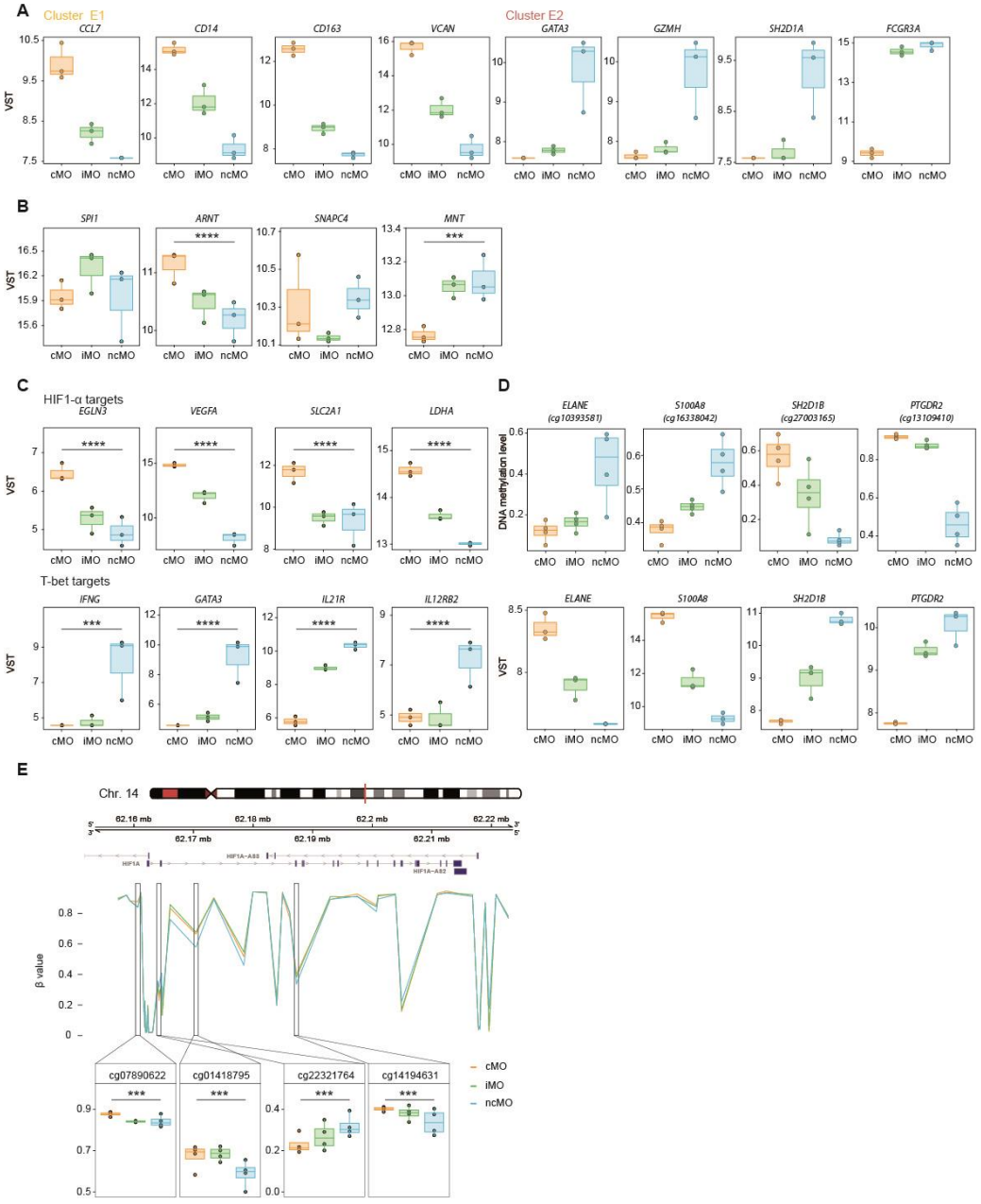
Finally, to inspect the impact of STAT1 dysregulation, initially determined from the bulk methylome and transcriptomic analyses, we studied the STAT1 regulon in the single-cell RNA-seq dataset. As anticipated, we observed an increased pattern of activation of this pathway in SLE samples. This activation was particularly exacerbated in the MHC-II–mono cluster and less in the S100A–mono or the CD16+ clusters (CD16+–mono and C1Q–mono) (Figure 13J). Together, these results correlate with our previous findings of gene expression changes, indicating a higher STAT1 influence on the high disease activity samples, mainly in cMOs and iMOs, which are the samples with a heightened ratio of the MHC-II–mono cluster.

Supplementary Figure 1



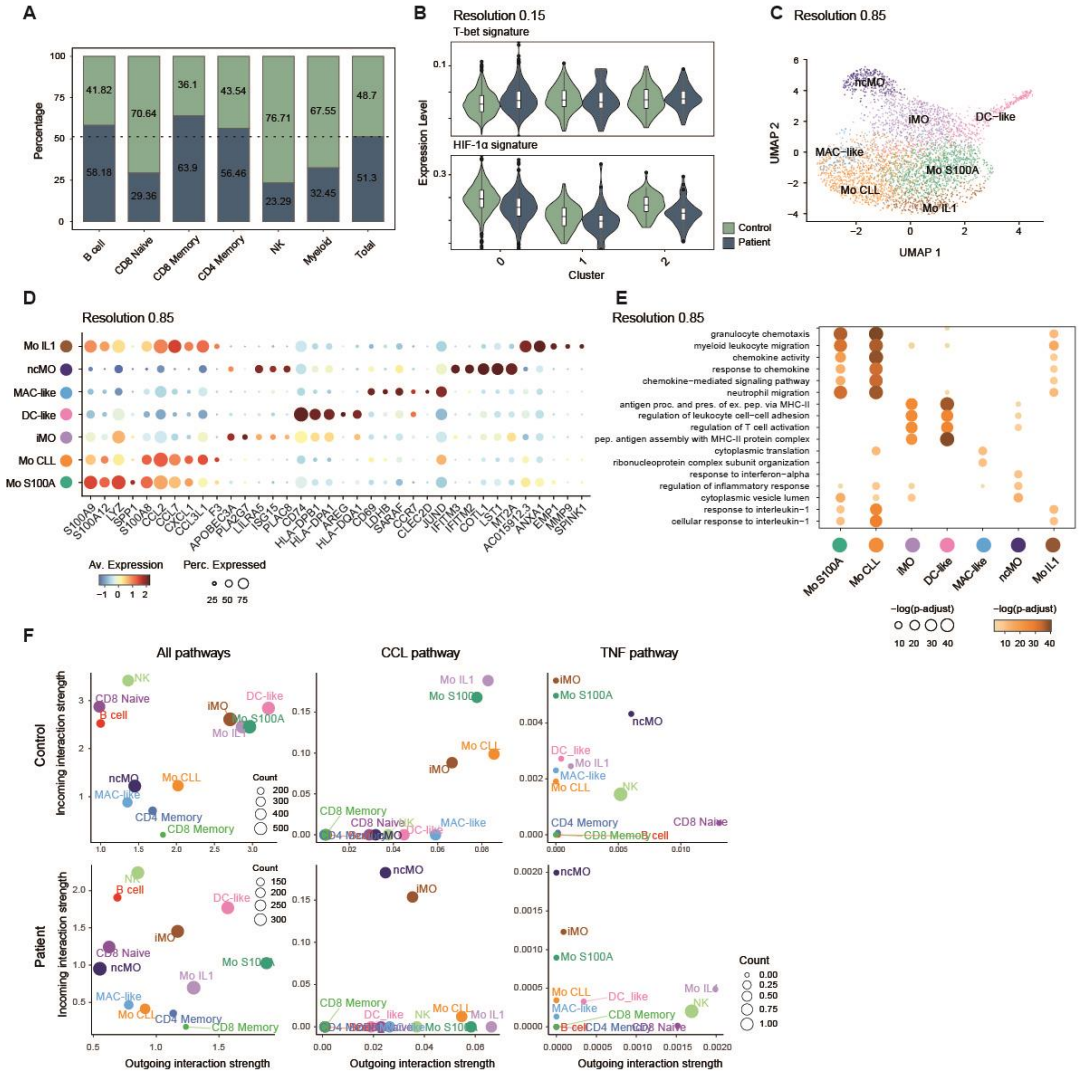
Supplementary Figure 1. **A**. Representation of the chromatin state annotations enriched in each cluster of DMPs. Significant is considered those enrichments with FDR < 0.01 and OR > 1.5. **B**. Boxplot of representative examples of the methylation value of four CpGs from each cluster of DMPs.

Supplementary Figure 2



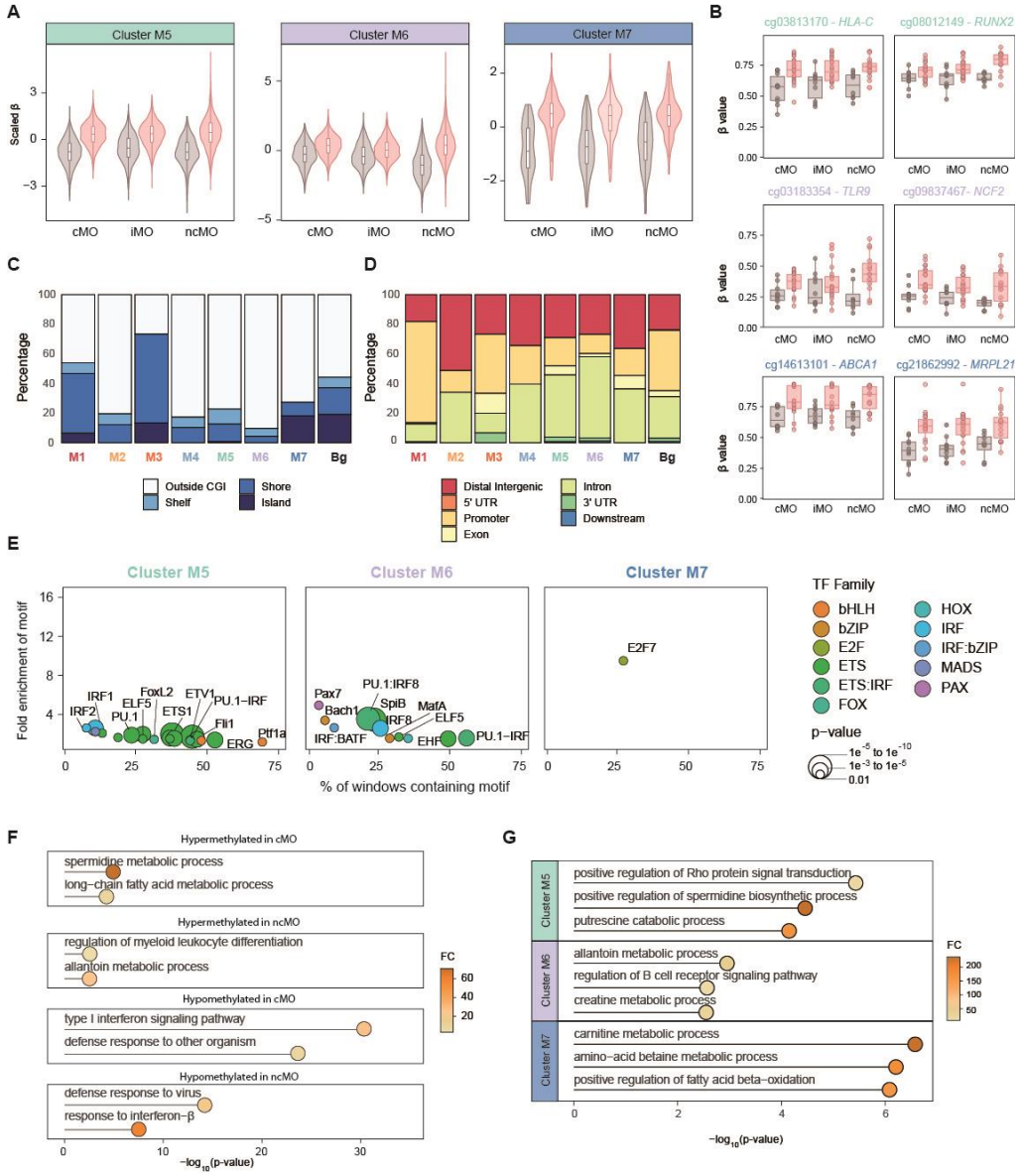
Supplementary Figure 2. **A**. Boxplot of representative examples of expression values of four genes from each cluster of DEGs. **B**. Boxplot representation of the top transcription factors predicted to be responsible for cMO and ncMO transcriptome. **C**. Boxplot representation of representative examples of target genes from HIF-1α (top) and T-bet (bottom) regulons. **D**. Boxplot of representative genes with correlation between the methylation of CpGs in the promoter of genes (top) and their expression (bottom). **E**. Graphical representation of the methylation status of the CpG surrounding *HIF-1α* gene.

Supplementary Figure 3

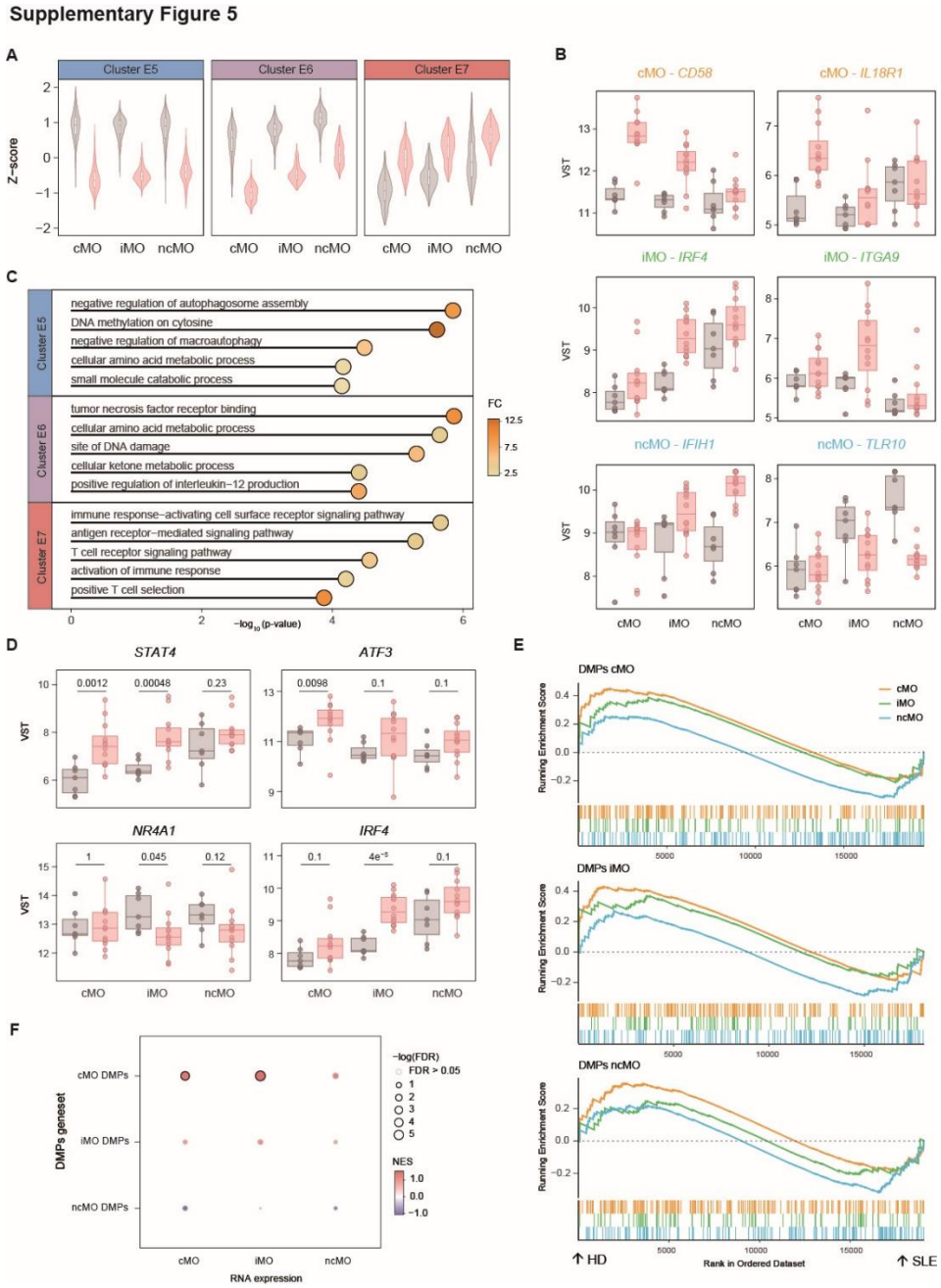


Supplementary Figure 3. **A**. Bar plot representation of the percentage contribution of cells from the affected individual (in blue) and his progenitor control (in green) to the cell compartments composing the PBMCs object of single-cell RNA-seq data EGAS00001004504. **B**. Violin plot representation of the HIF-1α and T-bet target genes in the clusters from the myeloid compartment as divided by a resolution of 0.15. **C**. Uniform Manifold Approximation and Projection (UMAP) of the myeloid compartment as divided by a resolution of 0.85. Annotations were done according to top marker genes and gene ontologies enriched in those. **D**. Representative genes among top differential markers between clusters of the myeloid compartment at a resolution of 0.85. **E**. Representative functions among those enriched within the top 50 marker genes of each cluster of the myeloid compartment at a resolution of 0.85. **F**. Scatter plot representation of the contribution of each group of cells to the cell-cell communications in the PBMC object of the control (top) and affected individual (bottom)

Supplementary Figure 4

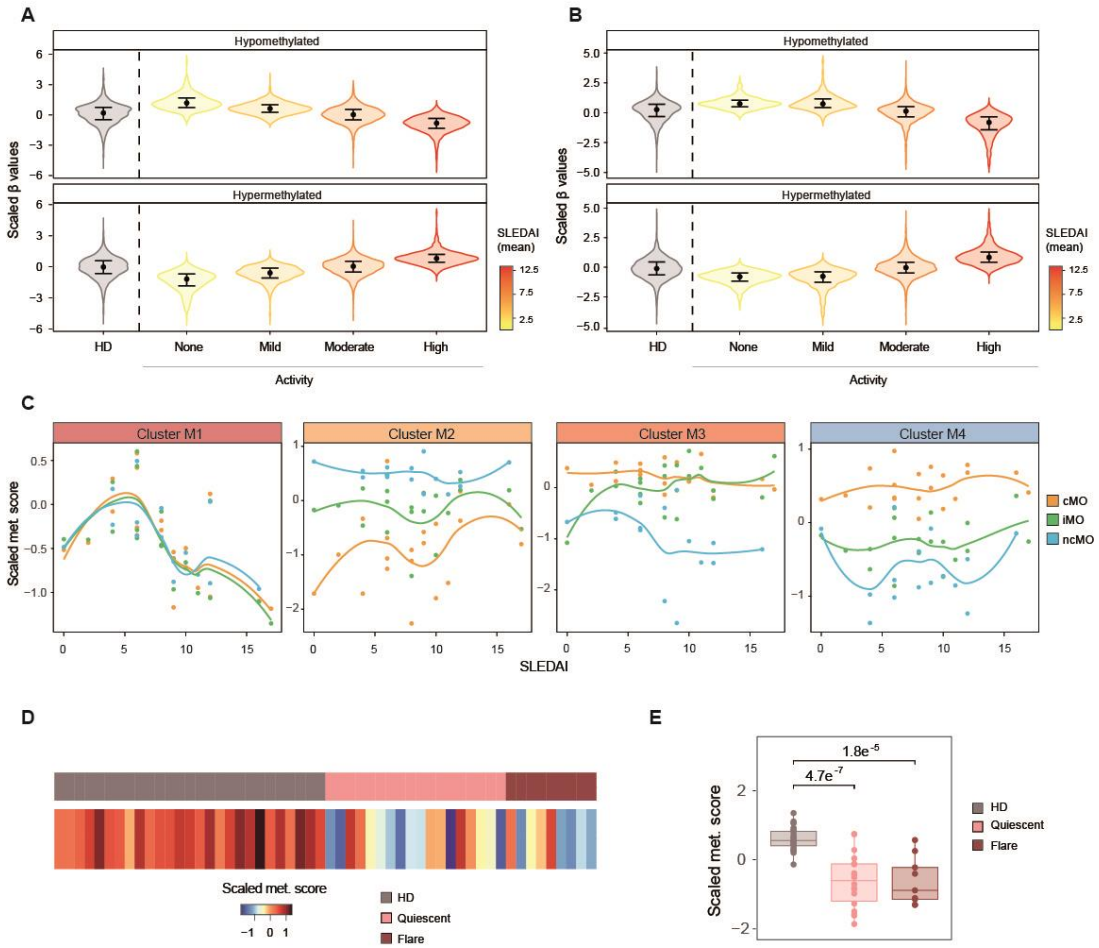


Supplementary Figure 4. A. Summary violin plot representation of the last three clusters of DMPs. In grey, HD samples, in pink, SLE. **B**. Illustrative examples of two individual DMPs from the last three clusters of DMPs. In grey, HD samples, in pink, SLE. The color of the CpG name represents the cluster it belongs to. **C** and **D**. Barplot representation of DMPs annotation per cluster to CpG Island context (C) and gene body context (D). **E**. Motif enrichment analysis of the sequence englobing the DMPs from cluster M5, M6, M7. **F**. Gene Ontology enrichment results of the genes annotated the closest to the DMPs from the comparisons HD vs SLE in cMO and ncMO, respectively. Hypomethylated refers to lower methylation levels in SLE and hypermethylation to higher methylation levels in SLE. **G**. Gene Ontology enrichment results of the genes annotated the closest to the DMPs from the last three clusters of DMPs.



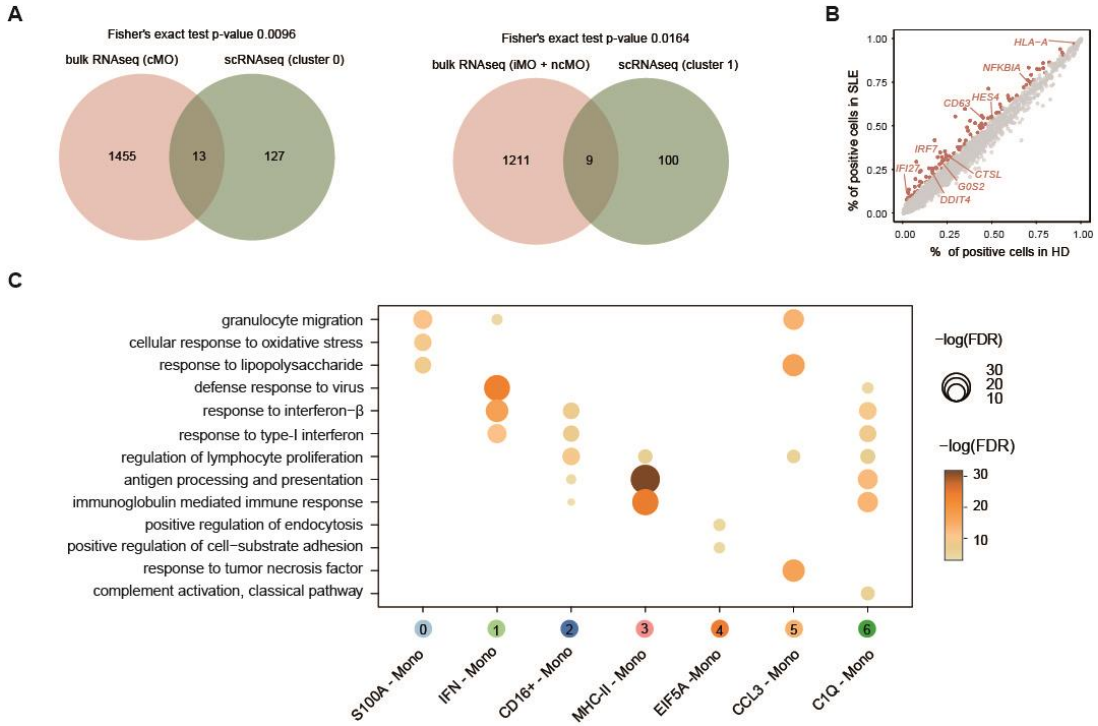
Supplementary Figure 5. A. Summary violin plot representation of the last three clusters of DEGs. In grey, HD samples, in pink, SLE. **B**. Illustrative examples of two DEGs from the comparison of HD vs SLE in each of the subsets. In grey, HD samples, in pink, SLE. **C**. Gene ontology enrichment of the last three clusters of DEGs. **D**. Expression of some highlighted transcription factors resulting from the motif enrichment analysis of DMPs. In grey, HD samples, in pink, SLE. **E**. Gene set enrichment analysis of the gene distribution between HD and SLE of genes annotating the closest to the DMPs in each subset. **F**. Summary of the gene set enrichment analysis statistical results.

Supplementary Figure 6



Supplementary Figure 6. A and **B**. Violin plot representation of the iMO's (A) and ncMO's (B) CpGs correlating with SLE activity as measured by SLEDAI in a Spearman's correlation. Samples divided in four groups according to their activity index: none is SLEDAI = 0, mild is SLEDAI between 1 and 5, moderate is an index between 6 and 10, and high is a SLEDAI above 10. **C**. Dotplot representation of methylation levels correlation with SLE activity index of the DMPs from clusters M1, M2, M3 and M4. **D**. Heatmap representation of STAT1 targets in a validation cohort of SLE and HD CD14+ monocytes (GSE59250). Methylation targets are defined per HOMER as in Figure 12C. **E**. Boxplot summary representation of STAT1 targets at the methylation level in a validation cohort of SLE and HD CD14+ monocytes (GSE59250).

Supplementary Figure 7



Supplementary Figure 7. A. Venn diagram representation of the overlap between the upregulated bulk DEGs (FDR < 0.05, log2FC > 1) and the single-cell DEGs (FDR < 0.05, log2FC > 0.1) in CD16- (left) and CD16+ (right) cells. **B**. Scatter plot representation of the correlation between the percentage of cells expressing each gene in the HD and SLE cohort in cluster 1 at the resolution of 0.05. In color the DEGs between SLE and HD samples in the single-cell dataset (FDR < 0.05, log2FC > 0.1). Labelled are the DEGs commonly upregulated in the bulk dataset comparison of HD versus SLE in either iMO or ncMO. **C**. Dotplot of representative gene ontology pathways enriched in the top 50 marker genes for each cluster generated at a resolution of 0.15.

5 DISCUSSION

In this thesis, we have investigated the phenotypes of the three main monocyte subsets: classical (cMOs), intermediate (iMOs) and non-classical (ncMOs) monocytes, both in physiological and pathological conditions. By combining the generation and integration of different omics datasets, we have identified individual factors and pathways that are important for the proper function of these cells.

In the two studies comprising this thesis, the results support that monocyte subsets undergo a sequential differentiation process with iMOs situated between the other two subsets. The process leading to the differentiation of the three subsets has been previously debated in the literature with two main hypotheses. The most accepted model proposes that there is a sequential differentiation from cMO to ncMO through iMO [20,269–271]. However, some work in murine monocytes challenged such model suggesting an independent origin for ncMOs directly from the bone marrow [272]. Nevertheless, this debate was conclusively resolved by two *in vivo* studies in humans that showed that the three monocyte subsets arise consecutively [27,273]. The only cell division takes place in the precursors from the bone marrow, which finally differentiate into cMOs. Once in the bloodstream, these may transition into iMOs and ncMOs, although these are not their most common fate. These results are in agreement with our data that shows a clear linear process with cMOs and ncMOs in the two extremes and with iMOs presenting an intermediate phenotype. This happens both at epigenetic and transcriptomic level in steady state and also in SLE environment.

The results of both studies are consistent with some of the previously assigned differential functions for each of the subsets and unveil novel phenotypic features. Although all three subsets were proven to be able to differentiate to monocyte-derived dendritic cells and macrophages by several groups [19,28,274], they appear to display preferences and different participation in most responses they partake. Even though it would be too simplistic to attribute a single function for each subset, we can better delineate their roles by defining cMOs in relation to bacterial response, iMOs by their implication in antigen-presenting functions and ncMOs with respect to viral

responses. Evidently, their full range of functions include a broader spectrum of processes, however, in relation to these three categories they have the ability to perform them dramatically different. As previously mentioned, cMOs have a high phagocytic capacity and high peroxidase activity, as well as secretion of high amounts of proinflammatory cytokines in response to LPS, as shown by functional *in vitro* assays [21,275,276]. In the case of iMOs, their increased capacity for presenting antigens is deduced in the literature from their unique transcriptomic profile with higher constitutive expression of genes from the MHC [20,277]. As for ncMOs, their attributed role as participants in the antiviral response derives from both transcriptomic profiles and results of *in vitro* stimulation with viral components [21]. These preferences are undoubtedly supported by the data generated in our two studies, through gene ontology analyses from either DMPs or DEGs. This in itself is an important result because, apart from the obvious relevance of the transcriptome in the functions attributed to a cell, we observe that these phenotypes are also imprinted at the epigenomic level. Thus, the identity of these cells is not just an accidental response to the environment but they are predisposed to it by their stable genomic structure.

In the two following sections, we discuss in detail the results obtained in each of the studies comprising this thesis.

5.1 PART 1. Epigenomic and transcriptomic changes of monocyte subsets differentiation

Our study demonstrates the occurrence of epigenomic and transcriptomic reprogramming in the differentiation from cMOs to ncMOs, which implicates an interplay between HIF-1α and T-bet. We have determined important and significant differences at both the epigenomic and transcriptomic level that correlate with the subset functions, supporting the participation of these changes in the acquired functions of ncMOs. Moreover, our results unveil the participation of an unexpected TF to have a role in the differentiation process, T-bet. This factor is known for being crucial in other cell types but has never been described in the myeloid compartment

until now. Finally, by inspecting scRNA-seq data from an individual with a mutation in T-bet, we have demonstrated that this TF contributes to proper monocyte differentiation.

The epigenetic profiles, in particular the DNA methylomes, of the monocyte subsets displayed a high number of significant differences across the differentiation process from cMO to ncMO. The majority of these differences were in the direction of loss of methylation. In general, iMO were found to be in an intermediate state between the other subsets, as has been observed for other characteristics of this intermediate subset [278,279]. Overall, these positions were strongly associated with immune related pathways. In particular, the hypomethylated positions in each subset associated with functions that had been previously related to said subset. For example, hypomethylated positions in cMO related with immune response to bacteria [21] while hypomethylated positions in ncMO associated with response and production of interferon signaling [19,30]. Interestingly, hypomethylated DMPs in cMO are associated with “positive regulation of fractalkine biosynthesis process”, a chemokine also known as CX3CL1 that is known to prevent apoptosis in all subsets [214]. Similarly, TF binding motifs encompassing these DMPs produced concordant results. The positions that lose methylation during the differentiation towards ncMO presented binding sequences for factors of the interferon regulatory factors family, suggesting a predisposition of this subset over the others to respond to these cytokines.

Transcriptionally, the subsets also presented a large number of differentially expressed genes which seconded our interest in their individual study. Again, the intermediate subset presented an in-between profile similar to that previously described [20]. Functions associated with increased expression in cMO related with their ascribed function as bacterium-response participants. Functions associated to increased ncMO expression pointed towards a regulation of T cell activation and differentiation, a role suggested for this subset by several groups [25,39].

We also found a difference in the TFs governing the transcriptional profiles of these differentiated cells. In particular, we have identified two transcription factors

that appear to have an intertwined role in the regulation of monocyte differentiation: HIF-1 and T-bet.

HIF-1 has been previously described to be critical for monocyte activation, particularly under stress conditions derived from hypoxic environments. In particular, HIF-1α regulates the metabolism, angiogenesis, invasion, cell survival, phagocytosis, secretion of cytokines and chemokine receptor levels of myeloid cells [280,281]. Thus, activation of this transcription factor is typically associated with enhanced activity of monocytes. In our study, we see higher levels of this transcription factors and its target genes in cMO in comparison to ncMO. We rationalize that the observed differences in this pathway activation might be due to the differences in the oxygen concentrations present in the bone marrow in comparison with those in the bloodstream. More explicitly, cMOs first appear in the bone marrow, where the oxygen concentration is far below normoxic conditions [219,220]. In the bloodstream, in normoxic conditions, they usually only live up to 24 hours, which might not be long enough for a representative general shift in this pathway expression. However, ncMOs come from iMOs and in total the two populations might spend around ten days in the bloodstream, a much reasonable time to adapt their gene expression regulatory mechanisms to normoxic conditions. Nevertheless, although the switch in HIF-1 expression and activity might respond to passive environmental changes, it does not discard that this transcription factor might have an active role in the transition from cMOs to ncMOs. One of this effects could directly be related to its influence in the expression of other transcription factors.

The second transcription factor shown to be relevant in the differentiation of monocyte subsets is T-bet, a transcription factor encoded by the gene *TBX21*. T-bet was first described as a master regulator for the commitment of T cells towards a Th1 phenotype [221]. Th1 cells are specialized in the elimination of virus and intracellular bacteria and exhibit a strongly proinflammatory phenotype through the production of IFN-γ and TNF-α [1]. Posteriorly, it was described that T-bet has an equally important role in the maturation of other immune cells such as NKs [282]. In NK cells, maturation

and effector functions are totally dependent on the proper expression of T-bet [283–285]. Finally, in myeloid cells, T-bet has only been described to be expressed in activated cells, upon IFN-γ stimulation [222,286]. Particularly, the induction of T-bet expression in mice dendritic cells was critical for their ability to activate a Th1 program upon interaction with naïve T cells, which could not take place with knock-out dendritic cells [286]. To the best of our knowledge, our results are the first to show a robust expression of T-bet in unstimulated myeloid cells, both at mRNA and at protein levels. Interestingly, this expression only takes place in iMOs and ncMOs, especially in the latter. This disparity could explain the discrepancy with previous results of unstimulated myeloid cells, where monocytes as a whole were analyzed and thus a majority of cMOs is to be expected in those studies. Focusing on this disparity between subsets, we studied the role the induction of T-bet could have in the differentiation process from cMO to ncMO. By using epigenomic and transcriptomic data of monocytes stimulated with different inflammatory cytokines, we inferred that that induction of T-bet expression by IFN-γ could explain part of the changes observed, in particular at epigenetic level. However, this stimulation could not recapitulate the changes observed at the transcriptomic level indicating that it is not the only pathway involved.

Finally, in order to better understand the impact of this factor in the myeloid compartment, we inspected the myeloid compartment of an individual with an homozygous loss-of-function mutation for T-bet, by making use of single-cell RNA-sequencing data from his peripheral blood [205]. This mutation caused him to develop Mendelian susceptibility to mycobacterial disease coursed by a deficient response to weakly virulent mycobacteria. In this patient, it had been exhaustively described the impact of this mutation in the CD4+ T cells and NK cells of the patient. Furthermore, our analyses revealed an unexpected generalized detriment in the myeloid compartment, with an aggravated effect on the ncMO cells. These cells had a clear altered responsive capacity to some important cytokine families such as CCL and TNF. Possible effects of the absence of TNF signals in monocytes are related with excessive monocyte death [287] and an aberrant activation of the inflammatory master

regulator NF-kβ [288]. Moreover, we detected an unexpected impact in HIF-1α signaling in this patient, supporting a potential regulatory axis between of these two factors. Interestingly, HIF-1α in monocytes has been shown to regulate the expression of several IRF family members including IRF5 and IRF3 [218]. IRF5 has parallelly been described to directly regulate the function of T-bet, at least in T and B cells [289,290]. Moreover, alterations in the components of this proposed axis are altered at multiple layers in the autoimmune disease of systemic lupus erythematosus [291].. An opposed behavior of HIF-1 and T-bet has also been observed in other cell types such as CD8+ T cells [292] and gut innate lymphoid cells [293]. Thus, it seems that the regulation of these TFs might be somehow related. The natural clearance of HIF-1 components in response to normal oxygen levels in the bloodstream might be one of the factors allowing the differentiation of cMO to ncMO, however further assays are required to validate this hypothesis.

5.2 PART 2. Epigenomic and transcriptomic changes in monocyte subsets in SLE

In this study, we exploited the power of multiomics to dissect changes in monocyte subpopulations in SLE. Previous studies had suggested a differential role for CD16+ monocytes through variations in their proportions and functions compared with those in healthy individuals [87–92]. Our study sheds light on both the differential phenotype of monocyte subsets in SLE and the expansion of certain subpopulations. On one hand, DNA methylation profiles revealed changes between SLE and HD individuals that are both common to all monocyte subsets and unique to one of the subpopulations. On the other hand, transcriptomic analyses showed that cMOs were the subset with higher and more significant differences between SLE and HD. These differences, both at epigenetic and transcriptomic levels, correlated with disease activity, particularly the STAT family of TFs. We also developed a model to predict the prognosis of the disease by inspecting the DNA methylation values of only three CpG sites. Finally, by analyzing a publicly available single-cell RNA-seq dataset of patients

with SLE, we deconvoluted the composition of our bulk data and identified a signature of cells that explain the phenotype observed in the SLE monocyte subsets.

The involvement of monocyte subsets in SLE and other inflammatory conditions has been subject of debate in recent years. Discrepancies in results in monocyte proportions are evident: some studies indicate higher percentage of cMOs [87], while others report an increase in iMOs [88,89] or in ncMOs [90,91]. Furthermore, certain studies have found no differences in the proportions of the subsets between SLE samples and HDs [92]. Our results indicate that the proportion of monocyte subsets varies in SLE in relation to disease activity. The proportion of iMOs is increased in SLE samples, particularly in those with low disease activity. The ncMO subset exhibits a similar trend, although statistical significance is not reached, likely due to cohort size limitations. These findings suggest a differential role of monocyte subsets at various stages of the disease.

DNA methylation changes can serve as indicators of environmental influence and can significantly affect cellular phenotype, predisposing cells to respond to subsequent stimuli. The chronic systemic inflammatory conditions characteristic of SLE contribute to shaping immune system cells into a state of constant responsiveness, perpetuating inflammation. Multiple studies have shown DNA methylation alterations in patients with SLE (reviewed in Ferreté-Bonastre et al [146]), generally pointing towards a dysregulated type I IFN signature. These modifications are associated with disease activity and prognosis. Our study on monocyte subsets has enabled us to identify epigenetic changes in positions shared among the different subsets, indicating a widespread inflammatory behavior primarily driven by this cytokine. However, we also observed DMPs exclusive to each monocyte subpopulation. Notably, among the hypomethylated DMPs unique to cMOs, up to 80% harbored the consensus sequence for binding the Fos/Jun family of TFs. In the myeloid compartment, these TFs and the demethylation of their target sequences are associated with differentiation to macrophages [294,295]. This suggests a priming of these cells towards macrophage differentiation, which would occur instead of

differentiation towards iMOs and ncMOs. In contrast, for DMPs exclusively demethylated in SLE ncMOs, we identified an association with T cell regulatory factors and pathways. Some of these DMPs were associated with the TF RORγt, a factor typically found in T cells, which has also been described in monocytes under some pathological conditions [239]. In this report, the expression of RORγt by a subset of monocytes was associated with their production of IL-17, a signature that resembles that of Th17 differentiation [296]. Interestingly, this type of T cell differentiation, particularly fostered by IFN-α-conditioned monocytes [297], has been shown to play a role in SLE pathogenesis [298,299]. Again, we see the implication of a transcription factor unexpected for this cell lineage that is activating a transcript profile similar to that expected for the cell lineage it was first discovered in. In both studies it has been a factor initially described for T cells that we have identified related with the ncMOs phenotype. This could point towards a close interaction between these two groups of cells, as previously described [36–38]. It also is a reflection of the plurifunctionality of transcription factors and cell processes under different environments.

It is possible to speculate that the observed differences in monocyte subsets could be due to the differential response because of the cells expressing different receptors, of them being primed differently or that the bulk samples are composed of a heterogeneous mixture of different cells. It is important to note that although these analyses yielded biologically relevant results, they were conducted with an FDR of 0.2, potentially increasing the number of false positives. To mitigate this, we applied a beta difference threshold filter of 10%; however, this still represents a limitation of our study.

The transcriptomic analysis revealed that alterations were predominantly shared among the three subsets, while subset-specific changes were primarily observed in cMOs. This subset displayed more and greater differences in gene expression compared with the other subsets. This finding likely suggests that cMOs are in a less differentiated state, making them more plastic and better able to respond to an

inflammatory environment. As anticipated, the genes upregulated in SLE were related to proinflammatory and antigen presentation pathways, whereas the downregulated genes were associated with negative regulation of autophagy or production of IL-12. These are relevant pathways associated with the development of SLE, where the dysregulated autophagosomic system may have an important role in the development and severity of the disease (reviewed in Liu et al [300]) and the IL-12/IL-23 axis is important for Th17 cell differentiation [297]. IFN signature, a well-known driver of SLE, was also present in our samples, particularly in those with higher activity as seen in both epigenetic and transcriptomic profile of STAT targets. In general terms, a panSTAT signature was observed in samples with high activity, a finding supported by multiple reports suggesting the prominent role of this family of TFs in SLE (reviewed in Goropevšek et al [301]). In particular, at the transcriptomic level of STAT1 targets, the member of the family more closely related to IFN signaling, it appears that cMOs and iMOs are the most responsive and susceptible subsets. In contrast, ncMOs' STAT1 targets expression does not reach a significant increase in high activity samples. Based on our previous study, we hypothesise that basal-state ncMOs already have an increased interferon signature associated with the gene expression changes linked to T-bet presence, and consequently interferon activation due to SLE is not as easy to identify. However, further studies would be required to validate this hypothesis.

Another transcriptomic feature exacerbated by disease activity in SLE is the difference in the expression levels of HIF-1α targets when comparing cMOs versus ncMOs. This signature was particularly elevated in cMOs from high activity samples but did not reach statistical significance in the comparison of the subsets within the HD samples. In SLE studies, HIF-1α expression has been associated with disease activity and renal affectation [302,303]. Importantly, inhibition of this factor in murine models improved several parameters of disease activity such as antibody levels, complement deposition in kidney or proteinuria [304]. In our samples, all monocyte subsets increase HIF-1α expression signature in SLE in comparison to HD, but this is particularly marked for the cMOs. As for T-bet expression signature, in all the

groups of samples, the expression was higher in ncMOs than in cMOs, which confirms the results observed in the first study conforming this thesis. However, we did not observe a difference between disease samples and healthy samples indicating that this TF is probably not affected by the pathological environment.

Based on the characteristics of our cohort and the clinical data collected regarding disease progression, we classified our patients into two groups: those with a good prognosis or in remission, and those with a poor prognosis or not in remission. Using this information, we developed a predictive model that used the epigenetic information from the first visit sample to predict the prognosis of the samples into these two categories. Remarkably, with the methylation status of just three CpG sites, we were able to achieve this classification in both our test and validation cohorts. This suggests that, even at the onset of a flare period, there are measurable indicators of the prognosis of the episode. A limitation of this aspect of our study is the size of our cohort, which necessitated division into independent groups for validation and testing. While these results should be further validated with a larger sample size, they are already promising, particularly considering the current sample size.

Finally, using a single-cell RNA-seq dataset from an SLE cohort, we identified subgroups of monocytes within our bulk dataset that help explain the differences observed. In this regard, cMOs and iMOs from SLE samples were formed by a high percentage of cells with an IFN signature, particularly in the high activity samples, which could explain the STAT1 signature that we had identified. Similarly, iMOs and ncMOs also contained a higher percentage of another group of cells with a high C1Q signature. This group of cells is likely to be important in SLE pathogenesis, where immunocomplexes formed by autoantibodies and complement factors are the main initiators of tissue damage. These results support the differential contribution of the monocyte subsets to SLE pathogenesis.

Most importantly, only a proportion of each sample was represented by these pathogenic cell phenotypes, but another relevant fraction was composed by cells with

a physiological phenotype. Thus, we observe a single-cell ability to adapt to the environment that causes some cells to have a pathological contribution while other are able to remain properly functional, evidencing the limitations of bulk studies in these cell types. Further investigation into the signals that influence a healthy monocyte to develop a proinflammatory, disease-promoting cell phenotype, or to follow a physiological destiny, even within the inflammatory environment present in this disease, would be of great interest.

6 CONCLUSIONS

1. The three monocyte subsets display differences in their DNA methylomes and transcriptomes in both physiological and pathological states.

2. The phenotypes inferred from the DNA methylome and transcriptome differences among monocyte subsets are consistent with some of their proposed functions.

3. HIF-1α and T-bet transcription programs are inversely correlated in the monocyte subsets, with the HIF-1α signature characterizing the cMOs' transcriptome and T-bet prevailing on ncMOs.

4. Induction of T-bet expression in cMOs can be accomplished by stimulation with IFN-γ but this exposure does not fully recapitulate the ncMO phenotype.

5. T-bet is required for the proper development of the myeloid compartment, especially for ncMOs.

6. In SLE, the three monocyte subsets display both common and subset specific DNA methylation alterations in relation to those from healthy donors.

7. In SLE, epigenomic and transcriptomic alterations in the three monocyte subsets are associated with alterations in interferon-signaling pathways.

8. Interferon-signaling pathway alterations are strongly correlated with disease activity in SLE, represented by changes in the expression and DNA methylation of STAT target genes.

9. cMO from SLE are predisposed towards a macrophage differentiation, as evidenced by the implication of specific transcription factors in their DNA methylome.

10. Single-cell RNA-seq analysis reveals a high heterogeneity in monocyte subsets, which varies between healthy donors and SLE patients, and is particularly marked in patients with high activity.

7 BIBLIOGRAPHY

1. Abbas AK, Lichtman AH, Shiv P. Cellular and molecular immunology. Jeremy Bowes; 2022.
2. Cooper MD, Alder MN. The evolution of adaptive immune systems. *Cell*. 2006;124(4):815–822.
3. Flajnik MF, Kasahara M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity*. 2001;15(3):351–362.
4. Pamcer Z, Amemiya CT, Ehrhardt GRA, et al. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature*. 2004;430(6996):174–180.
5. Liggett LA, Sankaran VG. Unraveling Hematopoiesis through the Lens of Genomics. *Cell*. 2020;182(6):1384–1400.
6. Guilliams M, Mildner A, Yona S. Developmental and Functional Heterogeneity of Monocytes. *Immunity*. 2018;49(4):595–613.
7. Morante-Palacios O, Fondelli F, Ballestar E, Martínez-Cáceres EM. Tolerogenic Dendritic Cells in Autoimmunity and Inflammatory Diseases. *Trends Immunol.* 2021;42(1):59–75.
8. Loughran SJ, Haas S, Wilkinson AC, Klein AM, Brand M. Lineage commitment of hematopoietic stem cells and progenitors: insights from recent single cell and lineage tracing technologies. *Exp. Hematol.* 2020;88:1–6.
9. Ginhoux F, Guilliams M. Tissue-Resident Macrophage Ontogeny and Homeostasis. *Immunity*. 2016;44(3):439–449.
10. Sierra-Filardi E, Puig-Kröger A, Blanco FJ, et al. Activin A skews macrophage polarization by promoting a proinflammatory phenotype and inhibiting the acquisition of anti-inflammatory macrophage markers. *Blood*. 2011;117(19):5092–5101.
11. Rodriguez RM, Suarez-Alvarez B, Lavín JL, et al. Signal Integration and Transcriptional Regulation of the Inflammatory Response Mediated by the GM-/M-CSF Signaling Axis in Human Monocytes. *Cell Rep.* 2019;29(4):860-872.e5.
12. Gaudino SJ, Kumar P. Cross-Talk Between Antigen Presenting Cells and T Cells Impacts Intestinal Homeostasis, Bacterial Infections, and

Tumorigenesis. *Front. Immunol.* 2019;10(MAR):

13. Goyert SM, Ferrero EM, Seremetis S V, et al. Biochemistry and expression of myelomonocytic antigens. *J. Immunol.* 1986;137(12):3909–3914.
14. Pugin J, Heumann D, Tomasz A, et al. CD14 is a pattern recognition receptor. *Immunity*. 1994;1(6):509–516.
15. Zanoni I, Granucci F. Role of CD14 in host protection against infections and in metabolism regulation. *Front. Cell. Infect. Microbiol.* 2013;3(JUL):
16. Zanoni I, Tan Y, Di Gioia M, Springstead JR, Kagan JC. By Capturing Inflammatory Lipids Released from Dying Cells, the Receptor CD14 Induces Inflammasome-Dependent Phagocyte Hyperactivation. *Immunity*. 2017;47(4):697-709.e3.
17. Yeap WH, Wong KL, Shimasaki N, et al. CD16 is indispensable for antibody-dependent cellular cytotoxicity by human monocytes. *Sci. Rep.* 2016;6:.
18. Patel AA, Zhang Y, Fullerton JN, et al. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* 2017;214(7):1913–1923.
19. Boyette LB, MacEdo C, Hadi K, et al. Phenotype, function, and differentiation potential of human monocyte subsets. *PLoS One*. 2017;12(4):.
20. Wong KL, Tai JJY, Wong WC, et al. Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets. *Blood*. 2011;118(5):.
21. Cros J, Cagnard N, Woollard K, et al. Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity*. 2010;33(3):375–386.
22. Smedman C, Ernemar T, Gudmundsdotter L, et al. FluoroSpot Analysis of TLR-Activated Monocytes Reveals Several Distinct Cytokine-Secreting Subpopulations. *Scand. J. Immunol.* 2012;75(2):249–258.
23. Zawada AM, Rogacev KS, Rotter B, et al. SuperSAGE evidence for CD14++CD16+ monocytes as a third monocyte subset. *Blood*. 2011;118(12):.
24. E G-G, R Z, H K, et al. Identification of a novel dendritic cell-like subset of CD64(+) / CD16(+) blood monocytes. *Eur. J. Immunol.* 2001;31(1):48–56.

25. Wong KL, Yeap WH, Tai JJY, et al. The three human monocyte subsets: implications for health and disease. *Immunol. Res.* 2012;53(1–3):41–57.
26. Skrzeczyńska-Moncznik J, Bzowska M, Loseke S, et al. Peripheral blood CD14high CD16+ monocytes are main producers of IL-10. *Scand. J. Immunol.* 2008;67(2):152–159.
27. Patel AA, Zhang Y, Fullerton JN, et al. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* 2017;214(7):1913–1923.
28. Frankenberger M, Hofer TPJ, Marei A, et al. Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint. *Eur. J. Immunol.* 2012;42(4):957–974.
29. Kapellos TS, Bonaguro L, Gemünd I, et al. Human Monocyte Subsets and Phenotypes in Major Chronic Inflammatory Diseases. *Front. Immunol.* 2019;10(AUG):
30. Kyogoku C, Smiljanovic B, Grün JR, et al. Cell-specific type I IFN signatures in autoimmunity and viral infection: what makes the difference? *PLoS One*. 2013;8(12):.
31. Belge K-U, Dayyani F, Horelt A, et al. The proinflammatory CD14+CD16+DR++ monocytes are a major source of TNF. *J. Immunol.* 2002;168(7):3536–3542.
32. Frankenberger M, Sternsdorf T, Pechumer H, Pforte A, Ziegler-Heitbrock HWL. Differential cytokine expression in human blood monocyte subpopulations: a polymerase chain reaction analysis. *Blood*. 1996;87(1):373–377.
33. Ancuta P, Liu KY, Misra V, et al. Transcriptional profiling reveals developmental relationship and distinct biological functions of CD16+ and CD16- monocyte subsets. *BMC Genomics*. 2009;10:403.
34. Auffray C, Fogg D, Garfa M, et al. Monitoring of blood vessels and tissues by a population of monocytes with patrolling behavior. *Science*. 2007;317(5838):666–670.
35. Carlin LM, Stamatiades EG, Auffray C, et al. Nr4a1-dependent Ly6C(low)

monocytes monitor endothelial cells and orchestrate their disposal. *Cell*. 2013;153(2):362–375.

36. Peng Y, Latchman Y, Elkon KB. Ly6C(low) monocytes differentiate into dendritic cells and cross-tolerize T cells through PDL-1. *J. Immunol.* 2009;182(5):2777–2785.
37. Qu C, Edwards EW, Tacke F, et al. Role of CCR8 and other chemokine pathways in the migration of monocyte-derived dendritic cells to lymph nodes. *J. Exp. Med.* 2004;200(10):1231–1241.
38. Ziegler-Heitbrock L. The CD14+ CD16+ blood monocytes: their role in infection and inflammation. *J. Leukoc. Biol.* 2007;81(3):584–592.
39. Randolph GJ, Sanchez-Schmitz G, Liebman RM, Schäkel K. The CD16(+) (FcgammaRIII(+)) subset of human monocytes preferentially becomes migratory dendritic cells in a model tissue setting. *J. Exp. Med.* 2002;196(4):517–527.
40. Walker JA, McKenzie ANJ. TH2 cell development and function. *Nat. Rev. Immunol.* 2018;18(2):121–133.
41. Seidler S, Zimmermann HW, Bartneck M, Trautwein C, Tacke F. Age-dependent alterations of monocyte subsets and monocyte-related chemokine pathways in healthy adults. *BMC Immunol.* 2010;11:.
42. Cao Y, Fan Y, Li F, et al. Phenotypic and functional alterations of monocyte subsets with aging. *Immun. Ageing*. 2022;19(1):.
43. Heimbeck I, Hofer TPJ, Eder C, et al. Standardized single-platform assay for human monocyte subpopulations: Lower CD14+CD16++ monocytes in females. *Cytometry. A*. 2010;77(9):823–830.
44. Grip O, Bredberg A, Lindgren S, Henriksson G. Increased subpopulations of CD16(+) and CD56(+) blood monocytes in patients with active Crohn's disease. *Inflamm. Bowel Dis.* 2007;13(5):566–572.
45. Urra X, Villamor N, Amaro S, et al. Monocyte subtypes predict clinical course and prognosis in human stroke. *J. Cereb. Blood Flow Metab.* 2009;29(5):994–1002.
46. Schauer D, Starlinger P, Reiter C, et al. Intermediate monocytes but not TIE2-

expressing monocytes are a sensitive diagnostic indicator for colorectal cancer. *PLoS One*. 2012;7(9):.

47. Figueroa-Vega N, Marín-Aragón CI, López-Aguilar I, et al. Analysis of the percentages of monocyte subsets and ILC2s, their relationships with metabolic variables and response to hypocaloric restriction in obesity. *PLoS One*. 2020;15(2):.
48. Sampath P, Moideen K, Ranganathan UD, Bethunaickan R. Monocyte Subsets: Phenotypes and Function in Tuberculosis Infection. *Front. Immunol.* 2018;9:.
49. Cooper DL, Martin SG, Robinson JI, et al. FcγRIIIa expression on monocytes in rheumatoid arthritis: role in immune-complex stimulated TNF production and non-response to methotrexate therapy. *PLoS One*. 2012;7(1):.
50. Bain CC, Scott CL, Uronen-Hansson H, et al. Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6Chi monocyte precursors. *Mucosal Immunol.* 2013;6(3):498–510.
51. Serbina N V., Salazar-Mather TP, Biron CA, Kuziel WA, Pamer EG. TNF/iNOS-producing dendritic cells mediate innate immune defense against bacterial infection. *Immunity*. 2003;19(1):59–70.
52. Guilliams M, Movahedi K, Bosschaerts T, et al. IL-10 dampens TNF/inducible nitric oxide synthase-producing dendritic cell-mediated pathogenicity during parasitic infection. *J. Immunol.* 2009;182(2):1107–1118.
53. Dunay IR, DaMatta RA, Fux B, et al. Gr1(+) inflammatory monocytes are required for mucosal resistance to the pathogen Toxoplasma gondii. *Immunity*. 2008;29(2):306–317.
54. De Trez C, Magez S, Akira S, et al. iNOS-producing inflammatory dendritic cells constitute the major infected cell type during the chronic Leishmania major infection phase of C57BL/6 resistant mice. *PLoS Pathog.* 2009;5(6):.
55. Fife BT, Huffnagle GB, Kuziel WA, Karpus WJ. CC chemokine receptor 2 is critical for induction of experimental autoimmune encephalomyelitis. *J. Exp. Med.* 2000;192(6):899–905.
56. Izikson L, Klein RS, Luster AD, Weiner HL. Targeting monocyte recruitment in

CNS autoimmune disease. *Clin. Immunol.* 2002;103(2):125–131.

57. Nahrendorf M, Swirski FK, Aikawa E, et al. The healing myocardium sequentially mobilizes two monocyte subsets with divergent and complementary functions. *J. Exp. Med.* 2007;204(12):3037–3047.
58. Graubardt N, Vugman M, Mouhadeb O, et al. Ly6Chi Monocytes and Their Macrophage Descendants Regulate Neutrophil Function and Clearance in Acetaminophen-Induced Liver Injury. *Front. Immunol.* 2017;8(JUN):
59. Ramachandran P, Pellicoro A, Vernon MA, et al. Differential Ly-6C expression identifies the recruited macrophage phenotype, which orchestrates the regression of murine liver fibrosis. *Proc. Natl. Acad. Sci. U. S. A.* 2012;109(46):.
60. London A, Itskovich E, Benhar I, et al. Neuroprotection and progenitor cell renewal in the injured adult murine retina requires healing monocyte-derived macrophages. *J. Exp. Med.* 2011;208(1):23–39.
61. Rose NR, Bona C. Defining criteria for autoimmune diseases (Witebsky's postulates revisited). *Immunol. Today*. 1993;14(9):426–430.
62. Criswell LA, Pfeiffer KA, Lum RF, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am. J. Hum. Genet.* 2005;76(4):561–571.
63. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat. Rev. Immunol.* 2018;18(5):325–339.
64. Pisetsky DS. Pathogenesis of autoimmune disease. *Nat. Rev. Nephrol.* 2023;19(8):509–524.
65. Jog NR, James JA. Epstein Barr Virus and Autoimmune Responses in Systemic Lupus Erythematosus. *Front. Immunol.* 2021;11:.
66. Klareskog L, Stolt P, Lundberg K, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum.* 2006;54(1):38–46.
67. Tiniakou E, Christopher-Stine L. Immune-mediated necrotizing myopathy

associated with statins: history and recent developments. *Curr. Opin. Rheumatol.* 2017;29(6):604–611.

68. Bakhriansyah M, Souverein PC, van den Hoogen MWF, de Boer A, Klungel OH. Risk of Nephrotic Syndrome for Non-Steroidal Anti-Inflammatory Drug Users. *Clin. J. Am. Soc. Nephrol.* 2019;14(9):1355–1362.
69. Tian J, Zhang D, Yao X, Huang Y, Lu Q. Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann. Rheum. Dis.* 2023;82(3):351–356.
70. Demas KL, Costenbader KH. Disparities in lupus care and outcomes. *Curr. Opin. Rheumatol.* 2009;21(2):102–109.
71. Kiriakidou M, Editors S, Cotton D, Taichman D, Williams S. Systemic lupus erythematosus. *Ann. Intern. Med.* 2013;159(7):.
72. Fanouriakis A, Tziolos N, Bertsias G, Boumpas DT. Update on the diagnosis and management of systemic lupus erythematosus. *Ann. Rheum. Dis.* 2021;80(1):14–25.
73. Bombardier C, Gladman DD, Urowitz MB, et al. Derivation of the sledai. A disease activity index for lupus patients. *Arthritis Rheum.* 1992;35(6):630–640.
74. Gladman DD, Ibañez D, Urowltz MB. Systemic lupus erythematosus disease activity index 2000. *J. Rheumatol.* 2002;29(2):288–291.
75. Parodis I, Emamikia S, Gomez A, et al. Clinical SLEDAI-2K zero may be a pragmatic outcome measure in SLE studies. *Expert Opin. Biol. Ther.* 2019;19(2):157–168.
76. Feld J, Isenberg D. Why and how should we measure disease activity and damage in lupus? *Press. Medicale*. 2014;43(6P2):.
77. Absher DM, Li X, Waite LL, et al. Genome-Wide DNA Methylation Analysis of Systemic Lupus Erythematosus Reveals Persistent Hypomethylation of Interferon Genes and Compositional Changes to CD4+ T-cell Populations. *PLoS Genet.* 2013;9(8):.
78. Ulff-Møller CJ, Asmar F, Liu Y, et al. Twin DNA Methylation Profiling Reveals Flare-Dependent Interferon Signature and B Cell Promoter Hypermethylation in Systemic Lupus Erythematosus. *Arthritis Rheumatol.* 2018;70(6):878–890.

79. Pisetsky DS, Lipsky PE. New insights into the role of antinuclear antibodies in systemic lupus erythematosus. *Nat. Rev. Rheumatol.* 2020;16(10):565–579.
80. Aringer M, Costenbader K, Daikh D, et al. 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. *Ann. Rheum. Dis.* 2019;78(9):1151–1159.
81. Aringer M, Costenbader K, Daikh D, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol.* 2019;71(9):1400–1412.
82. Wallace DJ, Stohl W, Furie RA, et al. A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. *Arthritis Rheum.* 2009;61(9):1168–1178.
83. Furie R, Petri M, Zamani O, et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum.* 2011;63(12):3918–3930.
84. Ma C, Xia Y, Yang Q, Zhao Y. The contribution of macrophages to systemic lupus erythematosus. *Clin. Immunol.* 2019;207:1–9.
85. Korman BD, Huang CC, Skamra C, et al. Inflammatory expression profiles in monocyte-to-macrophage differentiation in patients with systemic lupus erythematosus and relationship with atherosclerosis. *Arthritis Res. Ther.* 2014;16(4):.
86. Mills CD. M1 and M2 Macrophages: Oracles of Health and Disease. *Crit. Rev. Immunol.* 2012;32(6):463–488.
87. Wu Z, Zhang S, Zhao L, et al. Upregulation of CD16- monocyte subsets in systemic lupus erythematous patients. *Clin. Rheumatol.* 2017;36(10):2281–2287.
88. Zhu H, Hu F, Sun X, et al. CD16+ Monocyte Subset Was Enriched and Functionally Exacerbated in Driving T-Cell Activation and B-Cell Response in Systemic Lupus Erythematosus. *Front. Immunol.* 2016;7(NOV):
89. López P, Rodríguez-Carrio J, Martínez-Zapico A, et al. Low-density granulocytes and monocytes as biomarkers of cardiovascular risk in systemic

lupus erythematosus. *Rheumatology (Oxford).* 2020;59(7):1752–1764.

90. Mukherjee R, Kanti Barman P, Kumar Thatoi P, et al. Non-Classical monocytes display inflammatory features: Validation in Sepsis and Systemic Lupus Erythematous. *Sci. Rep.* 2015;5:.
91. Jiang W, Zhang L, Lang R, Li Z, Gilkeson G. Sex differences in monocyte activation in systemic lupus erythematosus (SLE). *PLoS One*. 2014;9(12):.
92. Li Y, Lee PY, Sobel ES, et al. Increased expression of FcgammaRI/CD64 on circulating monocytes parallels ongoing inflammation and nephritis in lupus. *Arthritis Res. Ther.* 2009;11(1):.
93. Postal M, Vivaldo JF, Fernandez-Ruiz R, et al. Type I interferon in the pathogenesis of systemic lupus erythematosus. *Curr. Opin. Immunol.* 2020;67:87–94.
94. Niewold TB, Swedler WI. Systemic lupus erythematosus arising during interferon-alpha therapy for cryoglobulinemic vasculitis associated with hepatitis C. *Clin. Rheumatol.* 2005;24(2):178–181.
95. RÖNNBLOM LE, ALM G V., ÖBERG KE. Possible induction of systemic lupus erythematosus by interferon-alpha treatment in a patient with a malignant carcinoid tumour. *J. Intern. Med.* 1990;227(3):207–210.
96. Capobianchi MR, Uleri E, Caglioti C, Dolei A. Type I IFN family members: similarity, differences and interaction. *Cytokine Growth Factor Rev.* 2015;26(2):103–111.
97. Platanitis E, Demiroz D, Schneller A, et al. A molecular switch from STAT2-IRF9 to ISGF3 underlies interferon-induced gene transcription. *Nat. Commun.* 2019;10(1):.
98. Nallar SC, Kalvakolanu D V. Interferons, signal transduction pathways, and the central nervous system. *J. Interferon Cytokine Res.* 2014;34(8):559–576.
99. Waddell SJ, Popper SJ, Rubins KH, et al. Dissecting interferon-induced transcriptional programs in human peripheral blood cells. *PLoS One*. 2010;5(3):.
100. Manivasagam S, Klein RS. Type III Interferons: Emerging Roles in Autoimmunity. *Front. Immunol.* 2021;12:.

101. Al-Janadi M, Al-Balla S, Al-Dalaan A, Raziuddin S. Cytokine profile in systemic lupus erythematosus, rheumatoid arthritis, and other rheumatic diseases. *J. Clin. Immunol.* 1993;13(1):58–67.
102. Paradowska-Gorycka A, Wajda A, Stypinska B, et al. Variety of endosomal TLRs and Interferons (IFN-α, IFN-β, IFN-γ) expression profiles in patients with SLE, SSc and MCTD. *Clin. Exp. Immunol.* 2021;204(1):49–63.
103. Liu W, Li M, Wang Z, Wang J. IFN- γ Mediates the Development of Systemic Lupus Erythematosus. *Biomed Res. Int.* 2020;2020:.
104. Saulescu I, Ionescu R, Opris-Belinski D. Interferon in systemic lupus erythematosus-A halfway between monogenic autoinflammatory and autoimmune disease. *Heliyon*. 2022;8(11):.
105. Lo MS. Insights Gained From the Study of Pediatric Systemic Lupus Erythematosus. *Front. Immunol.* 2018;9(JUN):
106. Webb R, Kelly JA, Somers EC, et al. Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients. *Ann. Rheum. Dis.* 2011;70(1):151–156.
107. Bin Joo Y, Lim J, Tsao BP, et al. Genetic variants in systemic lupus erythematosus susceptibility loci, XKR6 and GLT1D1 are associated with childhood-onset SLE in a Korean cohort. *Sci. Rep.* 2018;8(1):.
108. Webber D, Cao J, Dominguez D, et al. Association of systemic lupus erythematosus (SLE) genetic susceptibility loci with lupus nephritis in childhood-onset and adult-onset SLE. *Rheumatology (Oxford).* 2020;59(1):90–98.
109. Wang YF, Zhang Y, Lin Z, et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* 2021;12(1):.
110. Bentham J, Morris DL, Cunninghame Graham DS, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* 2015;47(12):1457–1464.
111. Kozyrev S V., Abelson AK, Wojcik J, et al. Functional variants in the B-cell

gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.* 2008;40(2):211–216.

112. Deafen D, Escalante A, Weinrib L, et al. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* 1992;35(3):311–318.
113. Brunner CM, Horwitz DA, Shann MK, Sturgill BA, Davis IV JS. Clinical and immunologic studies in identical twins discordant for systemic lupus erythematosus. *Am. J. Med.* 1973;55(2):249–254.
114. Bustabad S, González T, Trujillo E. Systemic lupus erythematosus in childhood. Concordance in a pair of monozygotic twins for anti-dsDNA antibodies and discordance for disease expression. *J. Rheumatol.* 1997;24(7):1450–1451.
115. Takvorian SU, Merola JF, Costenbader KH. Cigarette smoking, alcohol consumption and risk of systemic lupus erythematosus. *Lupus*. 2014;23(6):537–544.
116. Barbhaiya M, Lu B, Sparks JA, et al. Influence of Alcohol Consumption on the Risk of Systemic Lupus Erythematosus Among Women in the Nurses' Health Study Cohorts. *Arthritis Care Res. (Hoboken).* 2017;69(3):384–392.
117. Bernatsky S, Smargiassi A, Barnabe C, et al. Fine particulate air pollution and systemic autoimmune rheumatic disease in two Canadian provinces. *Environ. Res.* 2016;146:85–91.
118. Tedeschi SK, Barbhaiya M, Malspeis S, et al. Obesity and the risk of systemic lupus erythematosus among women in the Nurses' Health Studies. *Semin. Arthritis Rheum.* 2017;47(3):376–383.
119. Mu Q, Zhang H, Luo XM. SLE: Another Autoimmune Disorder Influenced by Microbes and Diet? *Front. Immunol.* 2015;6(NOV):
120. Draborg A, Izarzugaza JMG, Houen G. How compelling are the data for Epstein-Barr virus being a trigger for systemic lupus and other autoimmune diseases? *Curr. Opin. Rheumatol.* 2016;28(4):398–404.
121. Parks CG, D'Aloisio AA, Sandler DP. Early Life Factors Associated with Adult-Onset Systemic Lupus Erythematosus in Women. *Front. Immunol.*

2016;7(MAR):

122. Chaligné R, Heard E. X-chromosome inactivation in development and cancer. *FEBS Lett.* 2014;588(15):2514–2522.
123. Fang H, Disteche CM, Berletch JB. X Inactivation and Escape: Epigenetic and Structural Features. *Front. cell Dev. Biol.* 2019;7:.
124. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462(7271):315–322.
125. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* 2006;103(5):1412–1417.
126. Mohn F, Schübeler D. Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet.* 2009;25(3):129–136.
127. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin*. 2018;11(1):.
128. Ball MP, Li JB, Gao Y, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* 2009;27(4):361–368.
129. Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. *RNA*. 2018;24(10):1351–1362.
130. Kreibich E, Krebs AR. Relevance of DNA methylation at enhancers for the acquisition of cell identities. *FEBS Lett.* 2023;597(14):1805–1817.
131. Riggs AD. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* 1975;14(1):9–25.
132. Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*. 1992;71(5):865–873.
133. Sharif J, Muto M, Takebayashi SI, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*. 2007;450(7171):908–912.

134. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999;99(3):247–257.
135. Li D, Guo B, Wu H, Tan L, Lu Q. TET Family of Dioxygenases: Crucial Roles and Underlying Mechanisms. *Cytogenet. Genome Res.* 2015;146(3):171–180.
136. Müller N, Ponkkonen E, Carell T, Khobta A. Direct and Base Excision Repair-Mediated Regulation of a GC-Rich cis-Element in Response to 5-Formylcytosine and 5-Carboxycytosine. *Int. J. Mol. Sci.* 2021;22(20):.
137. Klug M, Schmidhofer S, Gebhard C, Andreesen R, Rehli M. 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. *Genome Biol.* 2013;14(5):.
138. Sardina JL, Collombet S, Tian T V., et al. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell*. 2018;23(5):727-741.e9.
139. Matuleviciute R, Cunha PP, Johnson RS, Foskolou IP. Oxygen regulation of TET enzymes. *FEBS J.* 2021;288(24):7143–7161.
140. Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* 1992;89(5):1827–1831.
141. Susan JCi, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 1994;22(15):2990–2997.
142. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. DNA methylation: bisulphite modification and analysis. *Nat. Protoc.* 2006;1(5):2353–2364.
143. Kaluscha S, Domcke S, Wirbelauer C, et al. Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* 2022;54(12):1895–1906.
144. Yin Y, Morgunova E, Jolma A, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017;356(6337):.
145. Vanzan L, Soldati H, Ythier V, et al. High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its

binding sites. *Nat. Commun.* 2021;12(1):.

146. Ferreté-Bonastre AG, Cortés-Hernández J, Ballestar E. What can we learn from DNA methylation studies in lupus? *Clin. Immunol.* 2022;234:.
147. Imgenberg-Kreuz J, Almlöf JC, Leonard D, et al. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann. Rheum. Dis.* 2018;77(5):736–743.
148. Bolin K, Imgenberg-Kreuz J, Leonard D, et al. Variants in BANK1 are associated with lupus nephritis of European ancestry. *Genes Immun.* 2021;22(3):194–202.
149. Ross MT, Grafham D V., Coffey AJ, et al. The DNA sequence of the human X chromosome. *Nature*. 2005;434(7031):325–337.
150. Sawalha AH, Wang L, Nadig A, et al. Sex-specific differences in the relationship between genetic susceptibility, T cell DNA demethylation and lupus flare severity. *J. Autoimmun.* 2012;38(2–3):J216–J222.
151. A H, G G, D P, et al. Overexpression of X-linked genes in T cells from women with lupus. *J. Autoimmun.* 2013;41:60–71.
152. Lu Q, Wu A, Tesmer L, et al. Demethylation of CD40LG on the Inactive X in T Cells from Women with Lupus. *J. Immunol.* 2007;179(9):6352–6358.
153. Coit P, Renauer P, Jeffries MA, et al. Renal involvement in lupus is characterized by unique DNA methylation changes in naïve CD4+ T cells. *J. Autoimmun.* 2015;61:29–35.
154. Zhao M, Liu S, Luo S, et al. DNA methylation and mRNA and microRNA expression of SLE CD4+ T cells correlate with disease phenotype. *J. Autoimmun.* 2014;54:127–136.
155. Mok A, Solomon O, Nayak RR, et al. Genome-wide profiling identifies associations between lupus nephritis and differential methylation of genes regulating tissue hypoxia and type 1 interferon responses. *Lupus Sci. Med.* 2016;3(1):.
156. Renauer P, Coit P, Jeffries MA, et al. DNA methylation patterns in naïve CD4+ T cells identify epigenetic susceptibility loci for malar rash and discoid rash in systemic lupus erythematosus. *Lupus Sci. Med.* 2015;2(1):.

157. Lanata CM, Paranjpe I, Nititham J, et al. A phenotypic and genomics approach in a multi-ethnic cohort to subtype systemic lupus erythematosus. *Nat. Commun.* 2019;10(1):.
158. Joseph S, George NI, Green-Knox B, et al. Epigenome-wide association study of peripheral blood mononuclear cells in systemic lupus erythematosus: Identifying DNA methylation signatures associated with interferon-related genes based on ethnicity and SLEDAI. *J. Autoimmun.* 2019;96:147–157.
159. Sunahori K, Juang Y-T, Kyttaris VC, Tsokos GC. Promoter hypomethylation results in increased expression of protein phosphatase 2A in T cells from patients with systemic lupus erythematosus. *J. Immunol.* 2011;186(7):4508–4517.
160. Hedrich CM, Rauen T, Apostolidis SA, et al. Stat3 promotes IL-10 expression in lupus T cells through trans-activation and chromatin remodeling. *Proc. Natl. Acad. Sci. U. S. A.* 2014;111(37):13457–13462.
161. Coit P, Jeffries M, Altorok N, et al. Genome-wide DNA methylation study suggests epigenetic accessibility andtranscriptional poising of interferon-regulated genes in naïve CD4+ T cellsfrom lupus patients. *J. Autoimmun.* 2013;43:78–84.
162. Coit P, Ortiz-Fernandez L, Lewis EE, et al. A longitudinal and transancestral analysis of DNA methylation patterns and disease activity in lupus patients. *JCI Insight*. 2020;5(22):.
163. Wang GS, Zhang M, Li XP, et al. Ultraviolet B exposure of peripheral blood mononuclear cells of patients with systemic lupus erythematosus inhibits DNA methylation. *Lupus*. 2009;18(12):1037–1044.
164. Peters FS, Peeters AMA, Hofland LJ, et al. Interferon-gamma DNA methylation is affected by mycophenolic acid but not by tacrolimus after T-cell activation. *Front. Immunol.* 2017;8(JUL):
165. Yang Y, Tang Q, Zhao M, et al. The effect of mycophenolic acid on epigenetic modifications in lupus CD4+T cells. *Clin. Immunol.* 2015;158(1):67–76.
166. Li TWH, Yang H, Peng H, et al. Effects of S-adenosylmethionine and methylthioadenosine on inflammation-induced colon cancer in mice.

Carcinogenesis. 2012;33(2):427–435.

167. Li YP, Liu Y, Strickland FM, Richardson B. Age-dependent decreases in DNA methyltransferase levels and low transmethylation micronutrient levels synergize to promote overexpression of genes implicated in autoimmunity and acute coronary syndromes. *Exp. Gerontol.* 2010;45(4):312–322.
168. Richardson B, Scheinbart L, Strahler J, et al. Evidence for impaired t cell dna methylation in systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Rheum.* 1990;33(11):1665–1673.
169. E C, J G, J M, et al. Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity. *J. Immunol.* 1988;140(7):2197–2200.
170. Richardson B. Effect of an inhibitor of DNA methylation on T cells. II. 5-Azacytidine induces self-reactivity in antigen-specific T4+ cells. *Hum. Immunol.* 1986;17(4):456–470.
171. Richardson B, Kahn L, Lovett EJ, Hudson J. Effect of an inhibitor of DNA methylation on T cells. I. 5-Azacytidine induces T4 expression on T8+ T cells. *J. Immunol.* 1986;137(1):35–39.
172. Harmon CE, Portanova JP. Drug-induced lupus: clinical and serological studies. *Clin. Rheum. Dis.* 1982;8(1):121–135.
173. Herrada AA, Escobedo N, Iruretagoyena M, et al. Innate Immune Cells' Contribution to Systemic Lupus Erythematosus. *Front. Immunol.* 2019;10(MAR):
174. Ziegler-Heitbrock L, Ancuta P, Crowe S, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood*. 2010;116(16):.
175. Haschka D, Tymoszuk P, Bsteh G, et al. Expansion of Neutrophils and Classical and Nonclassical Monocytes as a Hallmark in Relapsing-Remitting Multiple Sclerosis. *Front. Immunol.* 2020;11:.
176. Petri M, Buyon J, Kim M. Classification and definition of major flares in SLE clinical trials. *Lupus*. 1999;8(8):685–691.
177. Buyon JP, Petri MA, Kim MY, et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. *Ann. Intern. Med.* 2005;142(12 Pt 1):

178. Bibikova M, Lin Z, Zhou L, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* 2006;16(3):383–393.
179. Calle-Fabregat CD La, Rodríguez-Ubreva J, Cañete JD, Ballestar E. Designing Studies for Epigenetic Biomarker Development in Autoimmune Rheumatic Diseases. *Rheumatol. Immunol. Res.* 2022;3(3):103–110.
180. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 2016;11(9):1650–1667.
181. Zhang Y, Park C, Bennett C, Thornton M, Kim D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res.* 2021;31(7):1290–1295.
182. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
183. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010;
184. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 2016;34(5):525–7.
185. Morante-Palacios O, Ballestar E. shinyÉPICo: a graphical pipeline to analyze Illumina DNA methylation arrays. *Bioinformatics*. 2021;37(2):257–259.
186. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2017;33(4):558–560.
187. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24(13):1547–1548.
188. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
189. Wang Q, Li M, Wu T, et al. Exploring Epigenomic Datasets by ChIPseeker. *Curr. Protoc.* 2022;2(10):.
190. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*.

2015;31(14):2382–2383.

191. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics*. 2017;33(15):2381–2383.
192. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*. 2010;38(4):576–89.
193. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods*. 2012;9(3):215–216.
194. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* 2017;12(12):2478–2492.
195. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 2010;28(5):495–501.
196. Gu Z, Hübschmann D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics*. 2023;39(1):.
197. Martorell-Marugán J, González-Rumayor V, Carmona-Sáez P. mCSEA: detecting subtle differentially methylated regions. *Bioinformatics*. 2019;35(18):3257–3262.
198. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.
199. Hahne F, Ivanek R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol. Biol.* 2016;1418:335–351.
200. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287.
201. Carlson M. org.Hs.eg.db: Genome wide annotation for Human. R package version 3.8.2. 2019;
202. Alvarez MJ, Shen Y, Giorgi FM, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 2016;48(8):838–847.
203. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 2019;29(8):1363–1375.

204. Müller-Dott S, Tsirvouli E, Vazquez M, et al. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Res.* 2023;51(20):10934–10949.
205. Yang R, Mele F, Worley L, et al. Human T-bet Governs Innate and Innate-like Adaptive IFN-γ Immunity against Mycobacteria. *Cell*. 2020;183(7):1826-1847.e31.
206. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29.
207. Perez RK, Gordon MG, Subramaniam M, et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science*. 2022;376(6589):.
208. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience*. 2018;7(7):.
209. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* 2021;12(1):.
210. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 2014;32(4):381–386.
211. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*. 2017;14(10):979–982.
212. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566(7745):496–502.
213. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 2019;37(7):773–782.
214. White GE, McNeill E, Channon KM, Greaves DR. Fractalkine promotes human monocyte survival via a reduction in oxidative stress. *Arterioscler. Thromb. Vasc. Biol.* 2014;34(12):2554–2562.
215. Friedman AD. Transcriptional control of granulocyte and monocyte development. *Oncogene*. 2007;26(47):6816–6828.
216. Noura M, Morita K, Kiyose H, et al. Pivotal role of DPYSL2A in KLF4-mediated

monocytic differentiation of acute myeloid leukemia cells. *Sci. Rep.* 2020;10(1):.

217. Guilliams M, Tamura T. Decrypting DC development. *Nat. Immunol.* 2019;20(9):1090–1092.
218. Peng T, Du SY, Son M, Diamond B. HIF-1α is a negative regulator of interferon regulatory factors: Implications for interferon production by hypoxic monocytes. *Proc. Natl. Acad. Sci. U. S. A.* 2021;118(26):.
219. Chow DC, Wenning LA, Miller WM, Papoutsakis ET. Modeling pO(2) distributions in the bone marrow hematopoietic compartment. II. Modified Kroghian models. *Biophys. J.* 2001;81(2):685–696.
220. Hao Y, Cheng D, Ma Y, Zhou W, Wang Y. The relationship between oxygen concentration, reactive oxygen species and the biological characteristics of human bone marrow hematopoietic stem cells. *Transplant. Proc.* 2011;43(7):2755–2761.
221. Szabo SJ, Kim ST, Costa GL, et al. A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell*. 2000;100(6):655–669.
222. Lighvani AA, Frucht DM, Jankovic D, et al. T-bet is rapidly induced by interferon-gamma in lymphoid and myeloid cells. *Proc. Natl. Acad. Sci. U. S. A.* 2001;98(26):15137–15142.
223. Martinez FO, Gordon S, Locati M, Mantovani A. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. *J. Immunol.* 2006;177(10):7303–7311.
224. Law PP, Holland ML. DNA methylation at the crossroads of gene and environment interactions. *Essays Biochem.* 2019;63(6):717–726.
225. Huang C, Bi J. Expression Regulation and Function of T-Bet in NK Cells. *Front. Immunol.* 2021;12:.
226. Duckworth BC, Lafouresse F, Wimmer VC, et al. Effector and stem-like memory cell fates are imprinted in distinct lymph node niches directed by CXCR3 ligands. *Nat. Immunol.* 2021;22(4):434–448.
227. Joshi NS, Cui W, Chandele A, et al. Inflammation directs memory precursor

and short-lived effector CD8(+) T cell fates via the graded expression of T-bet transcription factor. *Immunity*. 2007;27(2):281–295.

228. Rius J, Guma M, Schachtrup C, et al. NF-kappaB links innate immunity to the hypoxic response through transcriptional regulation of HIF-1alpha. *Nature*. 2008;453(7196):807–811.
229. Schütze S, Wiegmann K, Machleidt T, Krönke M. TNF-induced activation of NF-kappa B. *Immunobiology*. 1995;193(2–4):193–203.
230. Liu T, Zhang L, Joo D, Sun SC. NF-κB signaling in inflammation. *Signal Transduct. Target. Ther.* 2017;2:.
231. Oikawa D, Gi M, Kosako H, et al. OTUD1 deubiquitinase regulates NF-κB- and KEAP1-mediated inflammatory responses and reactive oxygen species-associated cell death pathways. *Cell Death Dis.* 2022;13(8):.
232. Rochford I, Joshi JC, Rayees S, et al. Evidence for reprogramming of monocytes into reparative alveolar macrophages in vivo by targeting PDE4b. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 2021;321(4):L686–L702.
233. Kramer PR, Winger V, Reuben J. PI3K limits TNF-alpha production in CD16-activated monocytes. *Eur. J. Immunol.* 2009;39(2):561–570.
234. Guha M, Mackman N. The phosphatidylinositol 3-kinase-Akt pathway limits lipopolysaccharide activation of signaling pathways and expression of inflammatory mediators in human monocytic cells. *J. Biol. Chem.* 2002;277(35):32124–32132.
235. Gerszten RE, Friedrich EB, Matsui T, et al. Role of phosphoinositide 3-kinase in monocyte recruitment under flow conditions. *J. Biol. Chem.* 2001;276(29):26846–26851.
236. Nakamura T, Datta R, Kharbanda S, Kufe D. Regulation of jun and fos gene expression in human monocytes by the macrophage colony-stimulating factor. *Cell Growth Differ.* 1991;2(6):267–272.
237. Zenz R, Eferl R, Scheinecker C, et al. Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease. *Arthritis Res. Ther.* 2008;10(1):.
238. Castro G, Liu X, Ngo K, et al. RORγt and RORα signature genes in human Th17 cells. *PLoS One*. 2017;12(8):.

239. Lourda M, Olsson-Åkefeldt S, Gavhed D, et al. Detection of IL-17A-producing peripheral blood monocytes in Langerhans cell histiocytosis patients. *Clin. Immunol.* 2014;153(1):112–122.
240. Kim HA, Lee HS, Shin TH, et al. Polyamine patterns in plasma of patients with systemic lupus erythematosus and fever. *Lupus*. 2018;27(6):930–938.
241. dos Santos M, Veronese FV, Moresco RN. Uric acid and kidney damage in systemic lupus erythematosus. *Clin. Chim. Acta.* 2020;508:197–205.
242. Coit P, Yalavarthi S, Ognenovski M, et al. Epigenome profiling reveals significant DNA demethylation of interferon signature genes in lupus neutrophils. *J. Autoimmun.* 2015;58:59–66.
243. Miller S, Tsou PS, Coit P, et al. Hypomethylation of STAT1 and HLA-DRB1 is associated with type-I interferon-dependent HLA-DRB1 expression in lupus CD8+ T cells. *Ann. Rheum. Dis.* 2019;78(4):519–528.
244. Zhang Y, Liu Q, Yang S, Liao Q. CD58 Immunobiology at a Glance. *Front. Immunol.* 2021;12:.
245. Hoffmann JC, Räuker HJ, Krüger H, Bayer B, Zeidler H. Decreased levels of a soluble form of the human adhesion receptor CD58 (LFA-3) in sera and synovial fluids of patients with rheumatoid arthritis. *Clin. Exp. Rheumatol.* 1996;14(1):23–29.
246. De Jager PL, Baecher-Allan C, Maier LM, et al. The role of the CD58 locus in multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* 2009;106(13):5264–5269.
247. Mende R, Vincent FB, Kandane-Rathnayake R, et al. Analysis of Serum Interleukin (IL)-1β and IL-18 in Systemic Lupus Erythematosus. *Front. Immunol.* 2018;9(JUN):
248. Gutzmer R, Langer K, Mommert S, et al. Human dendritic cells express the IL-18R and are chemoattracted to IL-18. *J. Immunol.* 2003;171(12):6363–6371.
249. Gauzzi MC, Purificato C, Conti L, et al. IRF-4 expression in the human myeloid lineage: up-regulation during dendritic cell differentiation and inhibition by 1alpha,25-dihydroxyvitamin D3. *J. Leukoc. Biol.* 2005;77(6):944–947.
250. Jin S, Song Y, Zhou L, et al. Depletion of CUL4B in macrophages ameliorates diabetic kidney disease via miR-194-5p/ITGA9 axis. *Cell Rep.* 2023;42(6):.

251. Van Eyck L, De Somer L, Pombal D, et al. Brief Report: IFIH1 Mutation Causes Systemic Lupus Erythematosus With Selective IgA Deficiency. *Arthritis Rheumatol. (Hoboken, N.J.)*. 2015;67(6):1592–1597.
252. Hess NJ, Felicelli C, Grage J, Tapping RI. TLR10 suppresses the activation and differentiation of monocytes with effects on DC-mediated adaptive immune responses. *J. Leukoc. Biol.* 2017;101(5):1245–1252.
253. Herrmann M, Voll RE, Zoller OM, et al. Impaired phagocytosis of apoptotic cell material by monocyte-derived macrophages from patients with systemic lupus erythematosus. *Arthritis Rheum.* 1998;41(7):1241–1250.
254. Huang X, Hua J, Shen N, Chen S. Dysregulated expression of interleukin-23 and interleukin-12 subunits in systemic lupus erythematosus patients. *Mod. Rheumatol.* 2007;17(3):220–223.
255. Qiu F, Song L, Yang N, Li X. Glucocorticoid downregulates expression of IL-12 family cytokines in systemic lupus erythematosus patients. *Lupus*. 2013;22(10):1011–1016.
256. Larosa M, Zen M, Gatto M, et al. IL-12 and IL-23/Th17 axis in systemic lupus erythematosus. *Exp. Biol. Med. (Maywood).* 2019;244(1):42–51.
257. Tang YY, Wang DC, Wang YQ, Huang AF, Xu WD. Emerging role of hypoxia-inducible factor-1α in inflammatory autoimmune diseases: A comprehensive review. *Front. Immunol.* 2023;13:.
258. Álvarez K, Villar-Vesga J, Ortiz-Reyes B, et al. Induction of NF-κB inflammatory pathway in monocytes by microparticles from patients with systemic lupus erythematosus. *Heliyon*. 2020;6(12):.
259. Brightbill HD, Suto E, Blaquiere N, et al. NF-κB inducing kinase is a therapeutic target for systemic lupus erythematosus. *Nat. Commun.* 2018;9(1):.
260. Barnabei L, Laplantine E, Mbongo W, Rieux-Laucat F, Weil R. NF-κB: At the Borders of Autoimmunity and Inflammation. *Front. Immunol.* 2021;12:.
261. Trizzino M, Zucco A, Deliard S, et al. EGR1 is a gatekeeper of inflammatory enhancers in human macrophages. *Sci. Adv.* 2021;7(3):.
262. de la Calle-Fabregat C, Rodríguez-Ubreva J, Ciudad L, et al. The synovial and blood monocyte DNA methylomes mirror prognosis, evolution, and treatment

in early arthritis. *JCI insight*. 2022;7(9):.

263. de la Calle-Fabregat C, Niemantsverdriet E, Cañete JD, et al. Prediction of the Progression of Undifferentiated Arthritis to Rheumatoid Arthritis Using DNA Methylation Profiling. *Arthritis Rheumatol. (Hoboken, N.J.)*. 2021;73(12):2229–2239.
264. Mok A, Rhead B, Holingue C, et al. Hypomethylation of CYP2E1 and DUSP22 Promoters Associated With Disease Activity and Erosive Disease Among Rheumatoid Arthritis Patients. *Arthritis Rheumatol.* 2018;70(4):528–536.
265. Zhao M, Zhou Y, Zhu B, et al. IFI44L promoter methylation as a blood biomarker for systemic lupus erythematosus. *Ann. Rheum. Dis.* 2016;75(11):1998–2006.
266. Lu MC, Lai NS, Chen HC, et al. Decreased microRNA(miR)-145 and increased miR-224 expression in T cells from patients with systemic lupus erythematosus involved in lupus immunopathogenesis. *Clin. Exp. Immunol.* 2013;171(1):91–99.
267. Karonitsch T, Feierl E, Steiner CW, et al. Activation of the interferon-gamma signaling pathway in systemic lupus erythematosus peripheral blood mononuclear cells. *Arthritis Rheum.* 2009;60(5):1463–1471.
268. Zheng W, Wang X, Liu J, et al. Single-cell analyses highlight the proinflammatory contribution of C1q-high monocytes to Behçet's disease. *Proc. Natl. Acad. Sci. U. S. A.* 2022;119(26):.
269. Geissmann F, Manz MG, Jung S, et al. Development of monocytes, macrophages, and dendritic cells. *Science*. 2010;327(5966):656–661.
270. Yona S, Jung S. Monocytes: subsets, origins, fates and functions. *Curr. Opin. Hematol.* 2010;17(1):53–59.
271. Hume DA, MacDonald KPA. Therapeutic applications of macrophage colony-stimulating factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling. *Blood*. 2012;119(8):1810–1820.
272. Hanna RN, Carlin LM, Hubbeling HG, et al. The transcription factor NR4A1 (Nur77) controls bone marrow differentiation and the survival of Ly6C- monocytes. *Nat. Immunol.* 2011;12(8):778–785.

273. Tak T, Van Groenendael R, Pickkers P, Koenderman L. Monocyte Subsets Are Differentially Lost from the Circulation during Acute Inflammation Induced by Human Experimental Endotoxemia. *J. Innate Immun.* 2017;9(5):464–474.
274. Menezes S, Melandri D, Anselmi G, et al. The Heterogeneity of Ly6Chi Monocytes Controls Their Differentiation into iNOS+ Macrophages or Monocyte-Derived Dendritic Cells. *Immunity*. 2016;45(6):1205–1218.
275. Zhao C, Zhang H, Wong WC, et al. Identification of novel functional differences in monocyte subsets using proteomic and transcriptomic methods. *J. Proteome Res.* 2009;8(8):4028–4038.
276. Anbazhagan K, Duroux-Richard I, Jorgensen C, Apparailly F. Transcriptomic network support distinct roles of classical and non-classical monocytes in human. *Int. Rev. Immunol.* 2014;33(6):470–489.
277. Lee J, Tam H, Adler L, et al. The MHC class II antigen presentation pathway in human monocytes differs by subset and is regulated by cytokines. *PLoS One*. 2017;12(8):.
278. Hijdra D, Vorselaars ADM, Grutters JC, Claessen AME, Rijkers GT. Phenotypic characterization of human intermediate monocytes. *Front. Immunol.* 2013;4:.
279. Ong SM, Teng K, Newell E, et al. A Novel, Five-Marker Alternative to CD16-CD14 Gating to Identify the Three Human Monocyte Subsets. *Front. Immunol.* 2019;10:1761.
280. Zarember KA, Malech HL. HIF-1alpha: a master regulator of innate host defenses? *J. Clin. Invest.* 2005;115(7):1702–1704.
281. Lewis JS, Lee JA, Underwood JCE, Harris AL, Lewis CE. Macrophage responses to hypoxia: relevance to disease mechanisms. *J. Leukoc. Biol.* 1999;66(6):889–900.
282. Klose CSN, Blatz K, d'Hargues Y, et al. The transcription factor T-bet is induced by IL-15 and thymic agonist selection and controls CD8αα(+) intraepithelial lymphocyte development. *Immunity*. 2014;41(2):230–243.
283. Townsend MJ, Weinmann AS, Matsuda JL, et al. T-bet regulates the terminal maturation and homeostasis of NK and Valpha14i NKT cells. *Immunity*.

2004;20(4):477–494.

284. Daussy C, Faure F, Mayol K, et al. T-bet and Eomes instruct the development of two distinct natural killer cell lineages in the liver and in the bone marrow. *J. Exp. Med.* 2014;211(3):563–577.
285. Knox JJ, Cosma GL, Betts MR, McLane LM. Characterization of T-bet and eomes in peripheral human immune cells. *Front. Immunol.* 2014;5(MAY):
286. Lugo-Villarino G, Maldonado-López R, Possemato R, Peñaranda C, Glimcher LH. T-bet is required for optimal production of IFN-gamma and antigen-specific T cell activation by dendritic cells. *Proc. Natl. Acad. Sci. U. S. A.* 2003;100(13):7749–7754.
287. Wolf Y, Shemer A, Polonsky M, et al. Autonomous TNF is critical for in vivo monocyte survival in steady state and inflammation. *J. Exp. Med.* 2017;214(4):905–917.
288. Mussbacher M, Derler M, Basílio J, Schmid JA. NF-κB in monocytes and macrophages - an inflammatory master regulator in multitalented immune cells. *Front. Immunol.* 2023;14:.
289. Brune Z, Rice MR, Barnes BJ. Potential T Cell-Intrinsic Regulatory Roles for IRF5 via Cytokine Modulation in T Helper Subset Differentiation and Function. *Front. Immunol.* 2020;11:.
290. Zou YR, Davidson A. Age-associated B cells acquire a new wrinkle. *Nat. Immunol.* 2018;19(4):317–318.
291. Manni M, Gupta S, Ricker E, et al. Regulation of age-associated B cells by IRF5 in systemic autoimmunity. *Nat. Immunol.* 2018;19(4):407–419.
292. Doedens AL, Phan AT, Stradner MH, et al. Hypoxia-inducible factors enhance the effector responses of CD8(+) T cells to persistent antigen. *Nat. Immunol.* 2013;14(11):1173–1182.
293. Krzywinska E, Sobecki M, Nagarajan S, et al. The transcription factor HIF-1α mediates plasticity of NKp46+ innate lymphoid cells in the gut. *J. Exp. Med.* 2022;219(2):.
294. Dekkers KF, Neele AE, Jukema JW, Heijmans BT, De Winther MPJ. Human monocyte-to-macrophage differentiation involves highly localized gain and

loss of DNA methylation at transcription factor binding sites. *Epigenetics Chromatin*. 2019;12(1):.

295. Zhao X, Bartholdy B, Yamamoto Y, et al. PU.1-c-Jun interaction is crucial for PU.1 function in myeloid development. *Commun. Biol.* 2022;5(1):.
296. Chong WP, Mattapallil MJ, Raychaudhuri K, et al. The Cytokine IL-17A Limits Th17 Pathogenicity via a Negative Feedback Loop Driven by Autocrine Induction of IL-24. *Immunity*. 2020;53(2):384-397.e5.
297. Santini SM, Lapenta C, Donati S, et al. Interferon-α-conditioned human monocytes combine a Th1-orienting attitude with the induction of autologous Th17 responses: role of IL-23 and IL-12. *PLoS One*. 2011;6(2):.
298. Talaat RM, Mohamed SF, Bassyouni IH, Raouf AA. Th1/Th2/Th17/Treg cytokine imbalance in systemic lupus erythematosus (SLE) patients: Correlation with disease activity. *Cytokine*. 2015;72(2):146–153.
299. Abou Ghanima AT, Elolemy GG, Ganeb SS, Abo Elazem AA, Abdelgawad ER. Role of T helper 17 cells in the pathogenesis of systemic lupus erythematosus. *Egypt. J. Immunol.* 2012;19(2):25–33.
300. Liu X, Qin H, Xu J. The role of autophagy in the pathogenesis of systemic lupus erythematosus. *Int. Immunopharmacol.* 2016;40:351–361.
301. Goropevšek A, Holcar M, Avčin T. The Role of STAT Signaling Pathways in the Pathogenesis of Systemic Lupus Erythematosus. *Clin. Rev. Allergy Immunol.* 2017;52(2):164–181.
302. Garchow B, Maque Acosta Y, Kiriakidou M. HIF-1α and miR-210 differential and lineage-specific expression in systemic lupus erythematosus. *Mol. Immunol.* 2021;133:128–134.
303. Deng W, Ren Y, Feng X, et al. Hypoxia inducible factor-1 alpha promotes mesangial cell proliferation in lupus nephritis. *Am. J. Nephrol.* 2014;40(6):507–515.
304. Zhao W, Wu C, Li LJ, et al. RNAi Silencing of HIF-1α Ameliorates Lupus Development in MRL/lpr Mice. *Inflammation*. 2018;41(5):1717–1730.

8 APPENDIX

1. **Ferreté-Bonastre AG**, Martínez-Gallo M, Morante-Palacios O, Calvillo CL, Calafell-Segura J, Rodríguez-Ubreva J, Esteller M, Cortés-Hernández J, Ballestar E. Disease activity drives divergent epigenetic and transcriptomic reprogramming of monocyte subpopulations in systemic lupus erythematosus. Ann Rheum Dis. 2024; ard-2023-225433.

2. **Ferreté-Bonastre AG**, Cortés-Hernández J, Ballestar E. What can we learn from DNA methylation studies in lupus? Clin Immunol. 2022; 234:108920.

3. Godoy-Tena G, Barmada A, Morante-Palacios O, de la Calle-Fabregat C, Martins-Ferreira R, **Ferreté-Bonastre AG**, Ciudad L, Ruiz-Sanmartín A, Martínez-Gallo M, Ferrer R, Ruiz-Rodriguez JC, Rodríguez-Ubreva J, Vento-Tormo R, Ballestar E. Epigenetic and transcriptomic reprogramming in monocytes of severe COVID-19 patients reflects alterations in myeloid differentiation and the influence of inflammatory cytokines. Genome Med. 2022; 14(1):134.

4. Català-Moll F*, **Ferreté-Bonastre AG***, Godoy-Tena G*, Morante-Palacios O, Ciudad L, Barberà L, Fondelli F, Martínez-Cáceres EM, Rodríguez-Ubreva J, Li T, Ballestar E. Vitamin D receptor, STAT3, and TET2 cooperate to establish tolerogenesis. Cell Rep. 2022; 38(3):110244.

5. Català-Moll F, **Ferreté-Bonastre AG***, Li T*, Weichenhan D, Lutsik P, Ciudad L, Álvarez-Prado ÁF, Rodríguez-Ubreva J, Klemann C, Speckmann C, Vilas-Zornoza A, Abolhassani H, Martínez-Gallo M, Dieli-Crimi R, Rivière JG, Martín-Nalda A, Colobran R, Soler-Palacín P, Kracker S, Hammarström L, Prosper F, Durandy A, Grimbacher B, Plass C, Ballestar E. Activation-induced deaminase is critical for the establishment of DNA methylation patterns prior to the germinal center reaction. Nucleic Acid Res. 2021; 49(9):5057-5073.

(*) These authors contributed equally to the work.

TRANSLATIONAL SCIENCE

Disease activity drives divergent epigenetic and transcriptomic reprogramming of monocyte subpopulations in systemic lupus erythematosus

Anna Guiomar Ferreté-Bonastre,[1] Mónica Martínez-Gallo,[2] Octavio Morante-Palacios,[3] Celia Lourdes Calvillo,[1] Josep Calafell-Segura,[1] Javier Rodríguez-Ubreva,[1] Manel Esteller,[4,5,6,7] Josefina Cortés-Hernández,[8] Esteban Ballestar [1,9]

Handling editor Josef S Smolen

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/ard-2023-225433).

For numbered affiliations see end of article.

Correspondence to
Dr Esteban Ballestar, Josep Carreras Research Institute, Badalona 08916, Spain; eballestar@carrerasresearch.org

Received 17 December 2023
Accepted 15 February 2024

To cite: Ferreté-Bonastre AG, Martínez-Gallo M, Morante-Palacios O, *et al*. *Ann Rheum Dis* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/ard-2023-225433

ABSTRACT

Objectives Systemic lupus erythematosus (SLE) is characterised by systemic inflammation involving various immune cell types. Monocytes, pivotal in promoting and regulating inflammation in SLE, differentiate from classic monocytes into intermediate and non-classic monocytes, assuming diverse roles and changing their proportions in inflammation. In this study, we investigated the epigenetic and transcriptomic profiles of these and novel monocyte subsets in SLE in relation to activity and progression.

Methods We obtained the DNA methylomes and transcriptomes of classic, intermediate, non-classic monocytes in patients with SLE (at first and follow-up visits) and healthy donors. We integrated these data with single-cell transcriptomics of SLE and healthy donors and interrogated their relationships with activity and progression.

Results In addition to shared DNA methylation and transcriptomic alterations associated with a strong interferon signature, we identified monocyte subset-specific alterations, especially in DNA methylation, which reflect an impact of SLE on monocyte differentiation. SLE classic monocytes exhibited a proinflammatory profile and were primed for macrophage differentiation. SLE non-classic monocytes displayed a T cell differentiation-related phenotype, with Th17-regulating features. Changes in monocyte proportions, DNA methylation and expression occurred in relation to disease activity and involved the STAT pathway. Integration of bulk with single-cell RNA sequencing datasets revealed disease activity-dependent expansion of SLE-specific monocyte subsets, further supported the interferon signature for classic monocytes, and associated intermediate and non-classic populations with exacerbated complement activation.

Conclusions Disease activity in SLE drives a subversion of the epigenome and transcriptome programme in monocyte differentiation, impacting the function of different subsets and allowing to generate predictive methods for activity and progression.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The proportions of monocyte subsets/subpopulations change in autoimmune/inflammatory diseases, including lupus. However, we do not know the implications of these changes and whether the phenotype changes over the course of these conditions.

WHAT THIS STUDY ADDS

⇒ We show a subset-specific epigenetic and transcriptomic reprogramming for classic, intermediate and non-classic monocytes in systemic lupus erythematosus (SLE), which deepens their distinctive function.

⇒ We demonstrate a relationship with disease activity and progression which involves the STAT1 pathway and allows us to generate predictive formulas.

⇒ We identify novel SLE-specific monocyte subsets which are associated with activity and show additional distinctive functions of these subsets in relation to disease.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Molecular markers of monocyte subsets-based DNA methylation might be used in the clinical setting to help determine disease activity and prognosis in SLE.

INTRODUCTION

Autoimmune diseases are chronic inflammatory conditions generally characterised by the presence of autoantibodies and autoreactive T cells. Most research efforts in autoimmune diseases have focused on B and T cells, as the adaptive immune system is pathologically dysregulated. However, inflammation also involves monocytes, macrophages and dendritic cells, which have recently become the focus of many autoimmune disease studies (reviewed in van Kempen *et al*[1]). These cells play a crucial role in the innate immune system, being primarily responsible for phagocytosis of pathogens, antigen presentation and production of cytokines.[2] Monocytes and their derived cells are altered in a wide range of autoimmune diseases, as shown either by their aberrant infiltration to tissues,[3 4] causing hyperactivation of adaptive immune system components,[5 6] by their presentation of autoantigens to T and B cells[7] or by their altered phagocytosis capacity.[8 9]

For years, monocytes were studied as a homogeneous population of immune cells. However,

the development of single-cell level techniques, including flow cytometry, revealed the wide heterogeneity that exists within monocytes. The most common classification of monocyte subpopulations is based on the presence of the surface proteins CD14 and CD16. CD14 serves as co-receptor for toll-like receptor 4 (TLR4) and mediates lipopolysaccharide (LPS) signalling, while CD16, also known as Fc gamma receptor IIIa, is involved in the removal of antigen–antibody complexes from the circulation, as well as other responses.[10 11] Based on the expression of these two surface proteins, human monocytes are classified in three groups: classic monocytes (cMOs, CD14+ CD16−), intermediate monocytes (iMOs, CD14+ CD16+) and non-classic monocytes (ncMOs, CD14^{dim} CD16+). Indeed, an *in vivo* study by Patel *et al*[12] convincingly demonstrated that monocyte subsets represent distinct stages of differentiation within the same cell lineage. According to their model, cMOs are released from the bone marrow into the bloodstream, where they reside for about 1 day before either extravasating to the tissues, undergoing cell death or differentiating into iMOs. The iMO subset can be found in the bloodstream for up to 4 days, during which the majority of them further differentiate into ncMOs, which can persist in the bloodstream for up to 7 days. The relevance of monocyte subsets has been underscored by reported changes in their proportions in many inflammatory conditions including multiple sclerosis,[13] atherosclerosis[14] and, more recently, in COVID-19.[15 16] This has also been described in systemic lupus erythematosus (SLE), although different studies report conflicting results.[17–22]

Apart from their differences in cell surface markers, monocyte subsets perform specific functions. In this regard, cMOs are often considered the most inflammatory subset based on their higher production of proinflammatory molecules and their propensity to differentiate to dendritic cells and macrophages. iMOs have been reported to perform both proinflammatory and anti-inflammatory functions in different contexts. Out of the three subsets, iMOs are the main producers of interleukin-10 (IL-10) upon TLR stimulation.[23] On the other hand, they also are the main monocyte subset that produces proinflammatory cytokines in a tumorous environment.[24] Finally, studies of ncMO functions have led to contradicting results: some describe them as the best producers of tumour necrosis factor alpha (TNF-α) upon TLR4 stimulation with LPS,[23 25] while others describe them as the least effective to produce proinflammatory cytokines.[26] Nevertheless, functions generally associated with ncMOs include maintenance of vascular homeostasis and indicate that they represent the first line of defence against viral pathogens.[27–29]

SLE is characterised by multiple organ inflammatory damages and a wide spectrum of autoantibodies. Both genetic predisposition and environmental factors contribute to its development. Genome-wide association studies have revealed many genetic variants conferring susceptibility to SLE, affecting different immune cell types, especially monocytes and other myeloid cells.[30] Particularly, genetic factors seem to have a more definite role in childhood-onset SLE, while environmental factors are thought to gain relevance for adult-onset SLE (reviewed in Charras *et al* and Barbhaiya and Costenbader[31 32]). Environmental factors can trigger changes in the phenotype via epigenetic mechanisms. The most studied epigenetic mark is DNA methylation. In this regard, many studies accumulated in recent years describe aberrant DNA methylation profiles in patients with SLE (reviewed in Ferreté-Bonastre *et al*[33]). Roughly, these studies report a general loss of DNA methylation in different immune cell types in patients with SLE, particularly associated with increased expression of interferon (IFN)-regulated genes. Type I IFN has a central role in the development and progression of SLE as widely demonstrated by studies showing genetic variance and overexpression of several effectors of these signalling pathways (reviewed in Postal *et al*[34]). DNA methylation alterations in SLE, and other autoimmune diseases, can both associate with changes in the expression of immune-related genes or serve as a sensor of alterations in cellular pathways, becoming useful biomarkers of gene dysregulation.[35]

To date, there are no epigenetic studies in SLE focusing on monocyte subsets, despite their central and specific roles in inflammation. In the present study, we set to inspect the potential existence of monocyte subset-specific epigenetic and transcriptional alterations in SLE, which could shed light on the implication of changes in monocyte subpopulations in the activity and progression of SLE.

METHODS

Sample collection

20 patients with SLE and 13 healthy donors (HDs) were included in the study. The HDs were matched with SLE donors based on age and sex. 14 out of 20 SLE donors were of Caucasian ethnicity and, similarly, the majority of HDs belonged to the same ethnic group. Participants gave both oral and written consent for their blood to be used for research purposes. All the patients with SLE fulfilled the 2019 European Alliance of Associations for Rheumatology/American College of Rheumatology classification criteria for SLE.[36] Samples from each patient with SLE were collected at two different time points, the first one at the onset of a new flare and the second one at the subsequent visit. For patients with SLE, the Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2k)[37] was registered at each extraction date. Definition of a flare episode was done with Safety of Estrogens in Lupus Erythematosus-SLEDAI Flare Index, which can be used with any version of SLEDAI.[38 39]

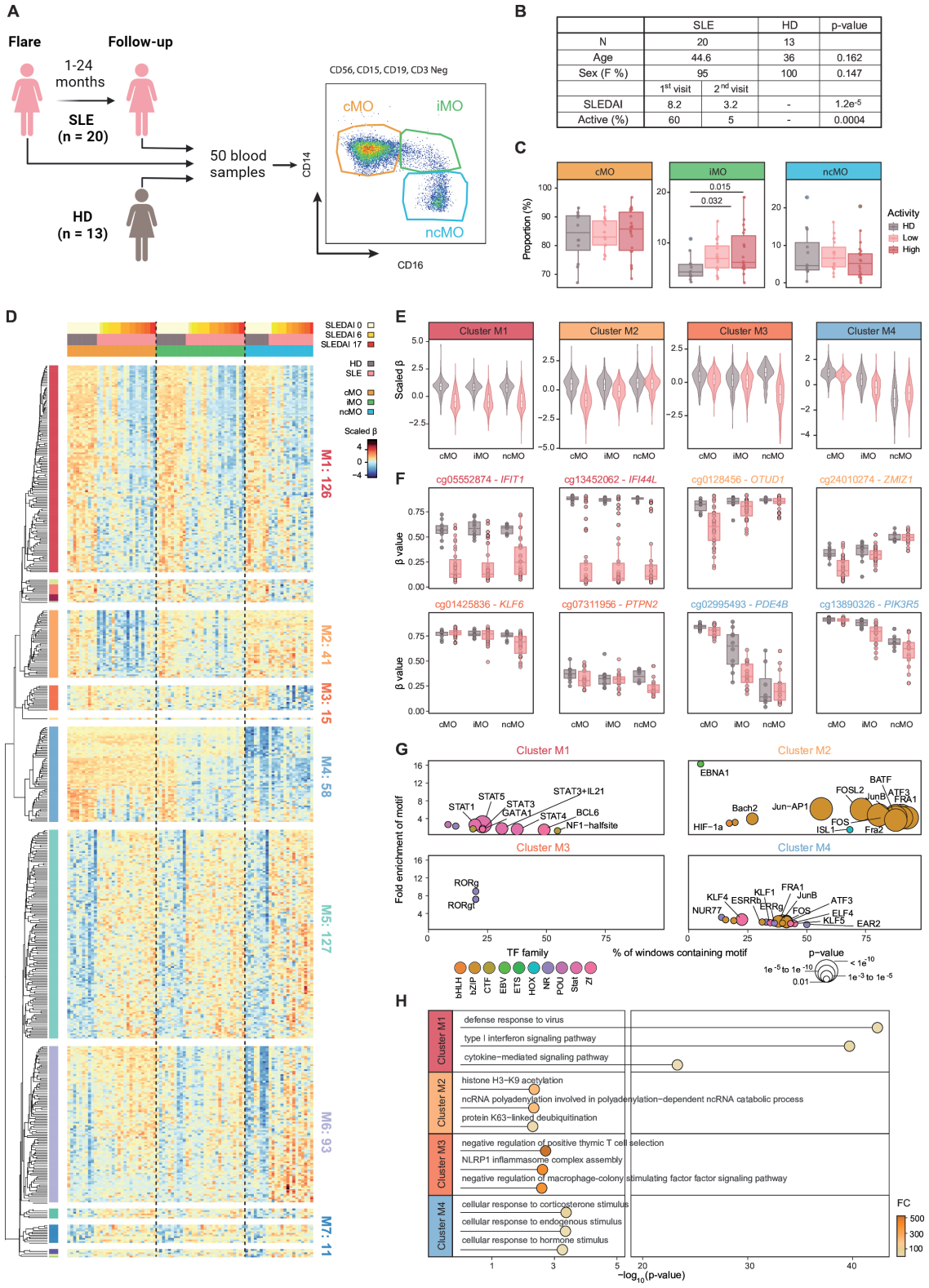
Sample processing and monocyte isolation

25 mL of whole blood were processed within 24 hours of collection by laying on Lymphocyte Separation Solution (Rafer, Zaragoza, Spain) and centrifuging without breaking in order to obtain peripheral blood mononuclear cells (PBMCs). Remaining erythrocytes were lysed with ACK lysis buffer. PBMCs were cryopreserved in fetal bovine serum with 10% dimethyl sulfoxide to gather several samples for flow cytometry sorting on the same batch. The day of cytometry sorting, samples were thawed at 37°C and stained for CD19-FITC (BD-Bioscience), CD15-FITC, CD3-FITC, CD56-PE, CD16-APC and CD14-APCVio 770 (Miltenyi). Finally, DAPI was included for the selection of viable cells. First, we filtered for singlets, FSC-SSC myeloid-like cells and DAPI-negative cells. Then a negative gating for CD15, CD3, CD19 and CD56 was performed as recommended in Mukherjee *et al* and Ziegler-Heitbrock *et al*[20 40] resulting in the obtention of the monocytic fraction. Finally, monocyte subsets were separated by positive selection based on the surface expression of CD14 and CD16 into cMO (CD14+ CD16−), iMO (CD14+ CD16+) and ncMO (CD14^{dim} CD16+) (figure 1A).

DNA and RNA extraction and preparation

Sorted samples were pelleted and frozen in RLT Buffer+10% β-mercaptoethanol as recommended by the kit Allprep DNA/RNA Micro, Mini (Qiagen). After collecting all the samples, double extraction of DNA and RNA was performed in as few batches as possible following manufacturer's instructions.

For DNA, the three monocyte subset samples from 20 patients in the first and second visits as well as 10 samples from



	SLE		HD	p-value
N	20		13	
Age	44.6		36	0.162
Sex (F %)	95		100	0.147
	1st visit	2nd visit		
SLEDAI	8.2	3.2	-	1.2e-5
Active (%)	60	5	-	0.0004

Figure 1 (A) Schematic representation of the cohort studied and the obtention of the monocyte subsets. (B) Summary table of the cohort characteristics. (C) Boxplot representation of the percentage of monocyte subsets within the monocyte fraction of PBMCs. Contains all samples collected, including second visits of SLE samples. Low activity are samples with SLEDAI of 5 or lower, high activity are samples with SLEDAI of 6 or higher. Statistics performed with two-tailed Wilcoxon test. (D) Heatmap representation of the 499 differentially methylated positions (DMPs) in either of the pairwise comparisons between SLE and HD samples from each subset, only first visit included. (E) Summary violin plot representation of the first four clusters of DMPs. In grey, HD samples; in pink, SLE. (F) Illustrative examples of two individual DMPs from the first four clusters. In grey, HD samples; in pink, SLE. The colour of the CpG name represents the cluster it belongs to. (G) Motif enrichment analysis of the sequence englobing the DMPs. (H) Gene Ontology enrichment results of the genes annotated closest to the DMPs from the first four clusters. cMO, classic monocyte; FC, fold change; HD, healthy donor; iMO, intermediate monocyte; ncMO, non-classic monocyte; PBMCs, peripheral blood mononuclear cells; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index; TF, transcription factor.

HDs were used; in total, 150 samples were analysed. 250 ng of genomic DNA were modified with bisulfite with EZ DNA methylation Gold kit (Zymo Research) following manufacturer's instructions. Modified material was hybridised in Infinium MethylationEPIC V.1.0 arrays, which interrogate around 850 000 CpG positions distributed throughout the genome. Since these arrays only accommodate eight samples each, a sample distribution strategy was devised to mitigate potential batch effects. Accordingly, disease state, subset, donor's age and sex, visit number, SLEDAI activity, as well as the time the sample spent frozen, were distributed in a balanced manner across arrays.[41]

For RNA, samples from the three monocyte subsets of 13 patients in the first visit, as well as seven samples from HDs were used. In total, 60 samples were sequenced in a 100 bp paired-end manner. Libraries and sequencing were performed by BGI Genomics (Hong Kong) with low-input transcriptome sequencing-Smart Seq-based method and DNBseq platform. Approximately 50 million reads were obtained per sample, and all samples passed sequencing quality control assessed with FastQC.[42]

DNA methylation analysis

Data from DNA methylation studies were analysed following the pipeline described for the shinyÉpico package.[43] In brief, after removing CpHs (where H=A, C or T), single nucleotide polimorphisms (SNPs) and X/Y chromosomes, samples were normalised using noob+quantile algorithms and beta values were transformed to M values with lumi package V.2.48.[44] Quality control was performed and sample composition was checked with the function *estimateCellCounts* from minfi package V.1.42[45]; some samples were removed due to undesired scores in either of the tests, all from the ncMO subset. In the end, 137 samples were retained for further analysis. At this stage, the association of potential covariates, including environmental factors and medications, with SLE samples was analysed. The absence of a significant association led to the exclusion of these factors from the analysis design. Statistical tests were performed with Limma package V.3.52[46] using the *arrayWeight* argument. Comparisons between SLE samples from the first visit and HDs were performed in each subset individually and CpGs with a difference in beta value >0.1 and a false discovery rate (FDR) <0.2 were considered differentially methylated positions (DMPs). Plots were generated with ggplot2 V.3.4 and gplots V.3.1 packages.

Annotation of CpGs to their closest gene was performed with the function *annotatePeak* from the package ChIPseeker V.1.32.[47 48] Motif enrichment analysis was performed with HOMER V.4.11[49] with the function *findMotifsGenome.pl*, within a window of 500 around the DMP. For this analysis, the arguments *–cpg* and *–nomotifs* were used and the background was considered as the total of CpGs included in the analysis. Gene Ontology analyses were performed with rGREAT package[50] with argument *rule* set to twoClosest genes and *version* of GREAT set to 4. Annotation to CpG island (CGI) context was performed with annotatr package V.1.22[51] and annotationHub V.3.4,[52] where shores are defined as 2 kb upstream/downstream of CpG islands, shelves are 2 kb further from shores and the rest is outside CGI. Annotation to genetic context was performed with ChIPseeker's function *annotatePeak*. Correlation of methylation and activity was performed with a Spearman's correlation with an estimate threshold of 0.7 and p value of <0.001.

Bulk RNA sequencing analysis

Files were aligned to hg38 and normalised with Kallisto.[53] Sample SLE15 was excluded from further analysis due to outlier profile. Differential analysis was performed with DESeq2[54] in each subset individually by comparing HD samples versus SLE samples. A batch effect was detected in some of the HD samples that were sorted and purified separately from the others including HD24, HD6, HD27 and HD28; this was then included as a covariate in posterior analyses. Differentially expressed genes (DEGs) were found by *lfcShrink* function with Ashr algorithm. DEGs were defined as protein-coding genes with an absolute $\log_2$ fold change ($\log_2$FC) >1 and an FDR <0.05. Variance stabilising transformation values and normalised counts provided by DESeq2 were used for visualisation purposes.

Plots were created with ggplot2 package.[55] Gene Ontology analyses were performed with *enrichGO* function from clusterProfiler package[56] with org.Hs.eg.db database.[57] Virtual Inference Protein-activity by Enriched Regulon Analysis (VIPER[58]) was performed with the dataset from the Collection of Transcriptional Regulatory Interactions[59] with the genes ranked by their p-adjust and $\log_2$FC. Gene Set Enrichment Analysis (GSEA) was performed with function *GSEA* from clusterProfiler[56] with the genes ranked by adjusted p value and $\log_2$FC. Gene sets from the DMPs were created by annotating the DMPs to their closest genes with *annotatePeak* from the ChIPseeker package.[47] STAT target scores were calculated with the function *run_viper* from the dorothea package.[60]

Single-cell RNA sequencing analysis and integration

For the single-cell RNA sequencing (RNA-seq) analysis and integration with the bulk RNA-seq dataset, we used the public dataset GSE174188.[61] The dataset was converted to a Seurat object with the function *Convert* from SeuratDisk. Posterior analyses of the data were performed with Seurat package V.4.3.0.[62] Good-quality cells were considered as those with a percentage of mitochondrial genes <15%, a number of counts >1500 and a number of features <3500 and >500. Analyses were performed exclusively on the myeloid fraction of the total object. After filtering, myeloid cells were re-integrated (with *k.weight* of 50), considering batch information. Doublets with T cells and platelets were removed. An object with 266k cells was obtained and used for clusterisation and annotation. For Uniform Manifold Approximation and Projection (UMAP) representation, the functions *ScaleData*, *RunPCA*, *FindNeighbors*, *FindClusters* and *RunUMAP* were used sequentially, with the first 30 dimensions. The resolution used is indicated in each figure. Given the fact that the ethnicities represented in the bulk RNA-seq dataset were primarily European and Hispanic, for the bulk integration analyses, these two cohorts were filtered resulting in a final object with 154k cells.

Clustering tree from the different resolutions was calculated and plotted with clustree package.[63] Top markers for each cluster were identified with the function *findAllMarkers* with the argument *min.cells.group*=50. Gene Ontology of top 50 markers was calculated with *enrichGO* function from clusterProfiler package[56] using as background all the genes included in the analysis. Pseudotime was calculated with the package monocle3.[64–66] Cell proportions from the bulk were measured with CIBERSORTx[67] and statistical inference of the proportions was conducted using a linear model with the function *lm* and a two-sided Wilcoxon rank-sum test.

RESULTS

Monocyte subset-specific DNA methylation profiles in SLE reveal a divergent shift in their epigenetic programmes

We collected peripheral blood samples from 20 patients with SLE during their initial hospital visits due to a flare episode of the disease or an increase in their symptoms (see the Methods section). Subsequently, we collected new blood samples from these same individuals during their follow-up visits to the doctor. These follow-up visits occurred within a period ranging from 1 to 24 months after the first visit, with 15 patients achieving remission. Additionally, we collected blood samples from 13 age-matched and sex-matched HDs (figure 1A,B). The clinical phenotype data are available in online supplemental table 1. Blood samples were processed for the isolation of the three monocyte subsets through a flow cytometry cell sorting strategy similar to previously published studies.[20 40] In brief, after obtaining the monocytic fraction by negative selection, monocyte subsets were separated based on the surface expression of CD14 and CD16 (figure 1A).

The percentages of the obtained monocyte subsets were analysed, revealing a significant increase in the proportion of iMOs in patients with SLE. Notably, this difference appeared to be more pronounced in patients with low disease activity (SLEDAI <6) compared with those with high activity (SLEDAI >6). The proportion of ncMOs also exhibited differential trends between these two groups of patients, once again showing relatively higher levels in patients with low disease activity (figure 1C). This finding was consistent with previous reports, which have indicated an increase in CD16+ monocytes. The association of these results with disease activity could help explain the disparity observed in previous reports, where some describe an increase in iMOs[18 19] while others reported an increase of ncMOs.[20 21]

We subsequently generated the DNA methylation profiles of cMOs, iMOs and ncMOs from both the SLE patient and HD cohorts. Our analysis revealed unique differences in DNA methylation patterns among the different monocyte subsets when comparing patients with SLE with HDs. Only the first visits of SLE samples were used in the upcoming analyses to avoid potential biases from biological replicates. Specifically, with an absolute differential of beta value >0.1 and an FDR <0.2, cMOs displayed 289 DMPs, iMOs exhibited 118 DMPs and ncMOs had 201 DMPs (figure 1D and online supplemental table 2). These produced a total of 499 unique DMPs, with some overlap between monocyte subsets. These DMPs were then grouped into seven clusters (M1–M7), based on the DNA methylation levels of the different samples. We identified clusters of DMPs exhibiting a similar behaviour across all monocyte subsets and clusters of DMPs displaying a distinctive profile in one of the subpopulations (figure 1E and online supplemental figure 1A). For instance, cluster M1 consists of 126 DMPs that are hypomethylated in SLE compared with HD across all three monocyte subpopulations. Examples of these DMPs include CpGs annotated to IFN-related genes, such as *IFIT1* and *IFI44L* (figure 1F). Cluster M2 (41 DMPs) annotate to genes that have a more drastic hypomethylation in cMOs than in both iMOs and ncMOs when comparing HDs versus SLE (figure 1D,E). It is the case of genes like *ZMIZ1*, a member of the protein inhibitor of activated STAT protein family, and *OTUD1,* a deubiquitinase related to TNF and IFN signalling[68] (figure 1F). Remarkably, cluster M3 (15 DMPs) corresponds to positions hypomethylated in SLE ncMOs in comparison with HD ncMOs (figure 1D,E). Examples include those annotating at genes like Krupple-like family transcription factor (TF) *KLF6*, which has been related to the aryl hydrocarbon receptor pathway, and *PTPN2* (figure 1F). Cluster M4 contains 58 DMPs that demethylate in the differentiation from cMOs to ncMOs and that in SLE iMOs are more advanced in the demethylation process than their HD counterparts (figure 1D,E). Examples of DMPs in the M4 cluster include those annotating to *PDE4B*, which encodes a key element for the monocyte's response to LPS[69] and *PIK3R5*, which encodes a regulatory subunit of PI3K complex, relevant for several immune functions in monocytes including cytokine release and adhesion[70–72] (figure 1F). The remaining clusters from M5 to M7 (127, 93 and 11 DMPs) were characterised by hypermethylation in SLE in comparison with HDs, being cluster M5 the predominant, with similar hypermethylation levels for all monocyte subsets (online supplemental figure 1A,B). The analysis of differentially methylated regions (DMRs) produced overlapping results with 60, 27 and 41 DMRs in cMOs, iMOs and ncMOs, respectively (data not shown). In summary, these results suggest that dysregulation of DNA methylation in SLE affects monocyte subsets in different ways. Since iMOs derive from cMOs and ncMOs from iMOs, one can interpret that some determinants in SLE pathology are affecting the differentiation process at the DNA methylation level.

Annotation of the DMPs to the genome in relation to CpG islands showed different patterns in various clusters (online supplemental figure 1C). For instance, clusters M1 and M3 showed an enrichment in CpGs annotated to shore regions in comparison with the background. In contrast, the remaining clusters presented a higher representation of positions outside CpG islands. In parallel, annotation of the DMPs in relation to the gene location also separated cluster M1 from the rest (online supplemental figure 1D). This showed a significant and marked representation of promoter regions encompassing more than 60% of the DMPs in this cluster. Cluster M3 also showed a high representation of promoter regions as well as exons. The remaining clusters showed a more homogeneous representation with similar region percentages to those present in the background. This highlights the common behaviour of the subsets in demethylating promoters, while subset-specific changes appear to participate in more complex regulatory processes.

TF-binding motif enrichment analysis in the regions surrounding DMPs showed that the different clusters exhibit significant enrichment for a wide range of TFs (figure 1G, online supplemental figure 1E and online supplemental table 3). For instance, DMPs from cluster M1 are enriched for binding motifs of several STAT TFs, indicating their relation to immune cell activation pathways. Additionally, approximately 80% of DMPs from cluster M2 (hypomethylated in cMOs from patients with SLE) presented a significant enrichment for TFs of the Fos and Jun family. These have been associated with monocyte-to-macrophage differentiation and activation,[73] as well as with other inflammatory diseases such as rheumatoid arthritis (reviewed in Zenz *et al*[74]). We were particularly interested in DMPs from cluster M3 because they are hypomethylated in SLE ncMOs, less studied than cMOs. DMPs from cluster M3 presented an enrichment for the consensus sequence of RORγ, which is a TF involved in Th17 differentiation.[75] This factor has also been described in a subset of monocytes which is associated with IL-17 production in pathological conditions.[76]

We also performed Gene Ontology analysis of the genes associated with these DMPs. Hypermethylated DMPs in SLE in comparison with HDs in either cMOs or ncMOs were associated with metabolic synthesis and degradation of several compounds such as 'spermidine biosynthesis process' or 'allantoin metabolic process' (online supplemental figure 1F and online supplemental

table 4). Notably, the polyamine spermidine has been previously described to be present at decreased concentrations in the plasma of patients with SLE.[77] Also, allantoin is the product of uric acid non-enzymatic oxidation. Urate may be found at elevated levels in the serum of patients with active SLE where it facilitates the activation of inflammatory pathways, particularly in those with kidney damage.[78] On the other hand, DMPs that lose methylation in SLE compared with HDs in either cMOs or ncMOs were annotated to genes related to immune system activation. DMPs exhibiting hypomethylation in cMOs were strongly associated with 'type I IFN signalling pathways', a proinflammatory group of cytokines that are a key player in SLE development and pathogenesis. This is consistent with previous results describing significant hypomethylation in IFN pathways in SLE immune cells.[79–83]

We also conducted Gene Ontology enrichment analysis of these positions when clustered in an unsupervised manner as depicted in figure 1D (see figure 1H, online supplemental figure 1G and online supplemental table 5). The results showed that DMPs in cluster M1, which are hypomethylated in SLE in all three monocyte subpopulations, strongly associated with type I IFN response pathways consistent with our previous analysis. Notably, cluster M3, which exhibited significant hypomethylation in SLE within the ncMO subset, was associated with pathways typically associated with T cells and pathways related negatively to monocyte differentiation to macrophages. This highlights once more the distinct involvement of ncMOs in SLE.

Monocyte subset-specific transcriptomic alterations in SLE

We then performed RNA-seq analysis of 60 samples, comprising the three monocyte subsets of 13 patients with SLE during their first visit and 7 HDs. The analysis revealed 2805 DEGs in cMOs, 1916 in iMOs and 1287 in ncMOs (FDR <0.5 and $\log_2$FC >1 or <−1) (online supplemental table 6). In general, the majority of DEGs corresponded to genes with higher expression levels in SLE compared with HDs (figure 3A,B). Also, most changes presented a similar behaviour in the three monocyte subsets, although a high percentage of the total DEGs did not reach statistical significance in ncMOs (figure 2B). DEGs were divided into seven main clusters (E1–E7) in an unsupervised manner based on their expression levels (figure 2B,C and online supplemental figure 2A). DEGs in E1 (986 DEGs) and E2 (631 DEGs) corresponded to genes that showed increased expression in SLE versus HDs in all subsets, with a more pronounced effect in cMOs. Notably, genes in cluster E2 displayed a progressive decrease in expression during the differentiation from cMOs to ncMOs. In cluster E3 (151 DEGs), genes showed similar expression levels in the three monocyte subsets in HDs and a similar upregulation in the three subsets in patients with SLE. In cluster E4 (70 DEGs), genes were downregulated during differentiation from cMOs to ncMOs but had a particular increase in expression in ncMOs of SLE. Cluster E5 (971 DEGs) and E6 (375 DEGs) showed a loss of expression in the three subsets in SLE. Finally, cluster E7 (322 DEGs) showed a parallel increase in expression during the differentiation from cMOs to ncMOs in both HDs and SLE.

Some genes with a unique behaviour in one of the subsets are represented in online supplemental figure 2B. Among these, *CD58* and *IL18R1* displayed a marked upregulation in cMOs from SLE. CD58 is a molecule expressed in the cell surface of monocytes, playing a crucial role in immune synapsis formation with CD2-expressing T and natural killer cells.[84] Dysregulation of this axis, characterised by aberrant expression of *CD58*, has been previously linked with several autoimmune diseases including rheumatoid arthritis.[85 86] On the other hand, the proinflammatory cytokine IL-18 has increased levels in the serum of patients with SLE[87] and the expression of *IL18R* in myeloid cells has been linked with their ability to migrate and be recruited to the site of inflammation.[88] The TF *IRF4*, which we found specifically upregulated in SLE iMOs, is expressed in myeloid cells upon stimulation with IFN-β and it induces their activation and differentiation.[89] *ITGA9*, associated with infiltration and migration in macrophages,[90] was found to be especially upregulated in iMOs from patients with SLE, indicating a potential migratory phenotype. In the case of ncMOs, *IFIH1* and *TLR10* were found to have specific differential expression in this subset, indicating a proinflammatory phenotype. *IFIH1* expression leads to a strong inflammatory response involving IFNs, a pathway known to be dysregulated in SLE.[91] In the case of *TLR10*, its expression in monocytes is linked to a suppression of their activation capacity and their ability to activate T cells.[92] In this case, ncMOs from HD increased their *TLR10* expression in comparison with cMOs, but this increase was not present in the patients with SLE which makes them more prone to this interaction than in physiological conditions.

Gene Ontology analysis revealed enrichment of biological functions similar to those obtained for DNA methylation (figure 2D, online supplemental figure 2C and online supplemental table 7). Clusters of DEGs that gain expression (clusters E1–E4) are primarily associated with immune response functions. For clusters E1 and E2, particularly upregulated in cMOs, functions related to 'leucocyte migration' and 'response to chemokine' are prevalent. Also, pathways of response through TNF, IL-1 and NF-κB were highly represented among the genes from these clusters. IFN production and major histocompatibility complex I (MHC-I) signalling are predominant among the functions derived from genes from cluster E3. Cluster E4, similarly to cluster E2, showed functions related to TNF production and response to bacteria molecules. For the clusters of genes undergoing downregulation, we observed enrichment in functions related to autophagy (cluster E5), probably related to the deficit of clearance of autophagocytic residues in SLE.[8] In cluster E6, genes related to several metabolic pathways and to the production of IL-12, a proinflammatory cytokine produced by myeloid cells downregulated by treatment with corticosteroids,[93 94] could lead to a T cell switch towards Th17 population.[95]

TF involvement was inferred from the gene expression of their targets using VIPER.[58] The TFs with the highest predicted activity in SLE in each of the subsets are shown in figure 2E. These results were consistent with the observation that the overall transcriptomic profile in SLE is relatively similar among monocyte subsets. HIF1A is the most significantly enriched TF in each of the subsets. This TF has been tightly linked with TNF-α signalling and inflammatory autoimmune responses (reviewed in Tang *et al*[96]). SLE monocytes are in a highly activated state, as also indicated by the enrichment of NF-kB, STATs or EGR1 regulons. NF-kB can be activated through TLR signalling or by uptake of microparticles in SLE becoming a therapeutic target.[97–99] EGR1 has also been shown to be associated with inflammatory responses in monocytes.[100]

In parallel, we inspected whether the TFs whose binding motifs were enriched in the different DMP clusters (figure 1G and online supplemental figure 1C) were also differentially expressed between the populations. This is the case for several of the factors (figure 2F and online supplemental figure 2D). For instance, cluster M2 DMPs were clearly enriched for consensus sequences of the family of TFs Fos-Jun. FOSL1 is one of the key members of this family and is differentially expressed in

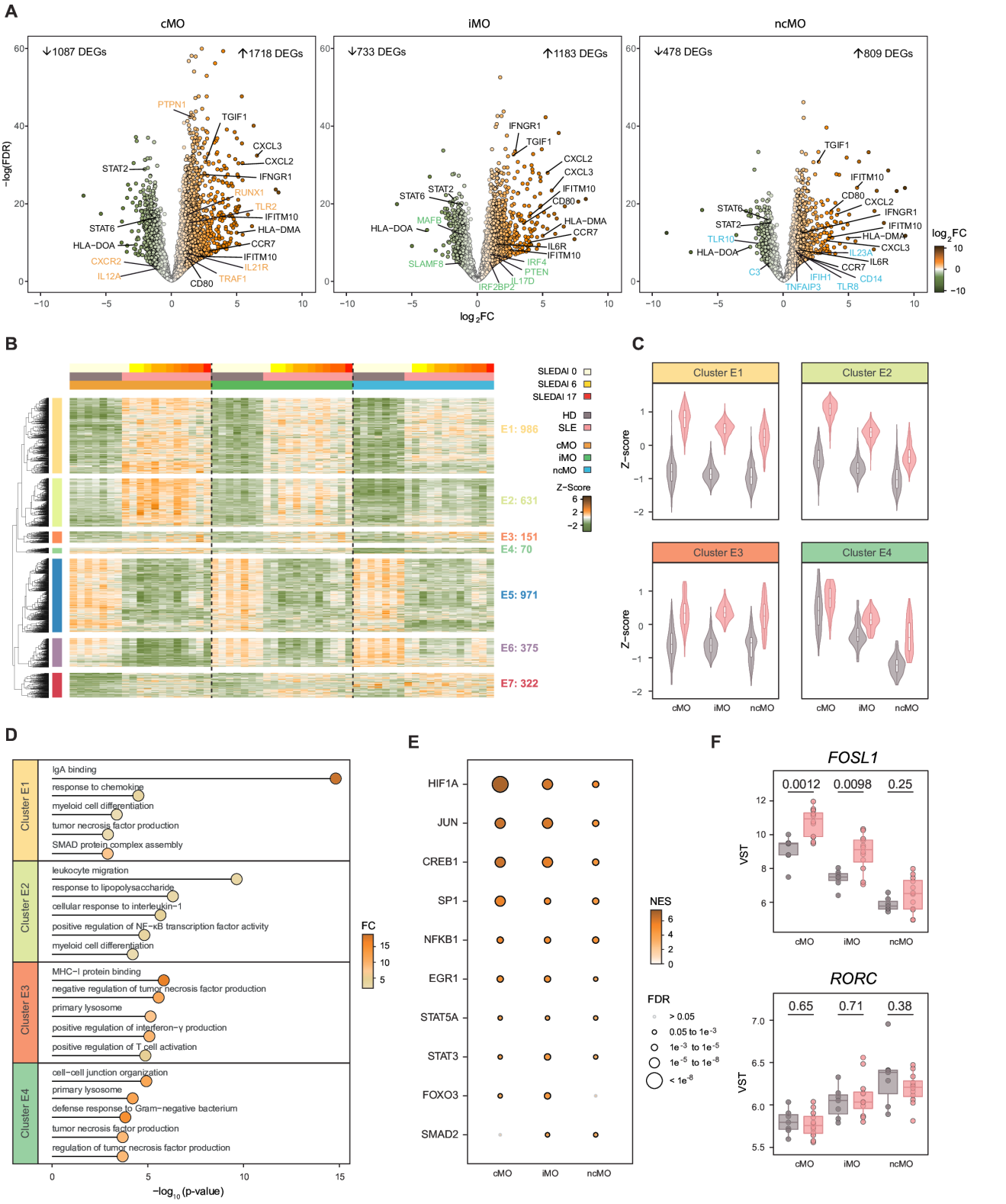


Figure 2 (A) Volcano plot of the differentially expressed genes (DEGs) in each pairwise comparison between SLE and HD from each subset of monocytes. Labels in black annotate highlighted DEGs common in the three comparisons. Labels in colour annotate highlighted DEGs unique to the respective subset. (B) Heatmap representation of the DEGs from the pairwise comparisons between SLE and HD in either of the subsets. (C) Summary violin plot representation of the first four clusters of DEGs. In grey, HD samples; in pink, SLE. (D) Gene Ontology enrichment of the first four clusters of DEGs. (E) Dotplot representation results of Virtual Inference of Protein-activity by Enriched Regulon analysis. It displays the top 10 transcription factors predicted to have more activity in SLE samples compared with HD samples, individually in each subset. (F) Expression of highlighted transcription factors resulting from the motif enrichment analysis of DMPs. In grey, HD samples; in pink, SLE. Statistics performed with two-tailed Wilcoxon test. cMO, classic monocyte; DMPs, differentially methylated positions; FDR, false discovery rate; HD, healthy donor; iMO, intermediate monocyte; log_2FC, log_2 fold change; ncMO, non-classic monocyte; NES, normalised enrichment score; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index; VST, variance stabilising transformation.

cMOs between HDs and SLE, but not in ncMOs, showing a similar behaviour to the DMPs from this cluster. In the case of cluster M3 of DMPs, where only one TF, RORγ, is significantly enriched in the sequences surrounding the CpG positions, its expression is not differential between HDs and SLE in any of the subpopulations. However, it has a clear upregulation between cMOs and ncMOs (figure 2F), suggesting a more prominent role in the transcriptome regulation of the latter. Other interesting TFs that present differences among the subpopulations include STAT4, ATF3, NR4A1 (*Nurr7*) and IRF4 (online supplemental figure 2D).

To further investigate the correlation between the epigenomic and transcriptomic profiles of the samples, we performed a GSEA of the genes annotated by proximity to the DMPs from each individual subpopulation (online supplemental figure 2E,F). The results revealed that the genes annotated to the DMPs from the cMO subset are significantly more enriched in transcriptomic profiles of cMOs from HD samples, in comparison with the SLE samples. Interestingly, cMO's DMPs were also enriched in the iMOs from HD samples, possibly due to the short lifespan of cMOs before they differentiate to iMOs which may cause cMO's epigenomic profile to influence iMO's transcriptomics. ncMO's DMPs did not reach statistical significance threshold for correlation, but they showed a tendency to be more enriched in the SLE samples.

DNA methylation changes correlate with SLE activity and progression

We and others have previously shown that the DNA methylation profiles of immune cells correlate with activity index in rheumatoid arthritis[101–103] and in SLE.[104] To test this potential relationship in the different monocyte subsets, we performed Spearman's correlation between the SLEDAI of the patients at their first visit and their DNA methylation profiles of the three monocyte subsets. We found 823 CpGs in cMOs, 683 in iMOs and 952 in ncMOs that correlated significantly with disease activity (ρ cut-off=0.7 and p value cut-off=1×10e–3) (figure 3A, online supplemental figure 3A,B and online supplemental table 8). These were vastly different between monocyte subsets. However, when analysing the Gene Ontology enrichment of the genes annotated to these positions, the results obtained were similar among the subsets. Particularly, the positions with a negative correlation to SLEDAI, that is, positions that are less methylated at higher activity indexes than at low activity indexes, associated with pathways related to immune response via type I IFN pathway in cMOs and iMOs (figure 3B and online supplemental table 9). These positions did not correlate with activity in the samples from the second visit, which contrasts with what we observed for other autoimmune conditions.[101] As a result, there were very poor correlations when comparing the difference in methylation with the difference in activity ($R^2<0.5$, data not shown).

Given the Gene Ontology enrichment results associating the correlating positions with IFN pathways, together with the potential role of STATs in the acquisition of the observed epigenetic profiles (figure 1G), we wondered whether these TFs could be associated with the activity in the samples. To address this question, we generated scores for each member of the STAT family for both DNA methylation and expression. For DNA methylation, we calculated the average of the scaled methylation values of all the CpGs with the corresponding STAT consensus sequence in a region of 500 bp surrounding the DMPs from cluster M1. For the expression data, we measured the average of the scaled expression values of the target genes of each STAT family member as defined by CollecTRI regulons.[59] The results revealed a significant positive association between the activity index and the levels of STAT targets (see figure 3C and online supplemental table 10). This association holds true for all members of the STAT family except STAT6, whose targets do not increase their expression along with SLE activity. We highlight the results for STAT1 given that it is the main responder to IFN signalling, and its expression is dysregulated in SLE.[105 106] Samples with higher activity decrease the methylation of STAT1-responsive DMPs while they have a higher expression of STAT1 gene targets (figure 4D,E). The correlation of these STAT1-associated DMPs with disease activity was validated with an external public cohort of CD14+ monocytes from 27 SLE donors and 27 HDs (GSE59250; see online supplemental figure 3D,E). Although the association between the epigenetic STAT1 profile and the activity occurs for the three subsets, the correlation of the transcriptomic profile is only statistically significant in cMOs and iMOs. This suggests that ncMOs are less STAT1 responsive.

Regarding the disease progression of the patients included in the study, 15 out of the 20 patients improved over the following months and achieved a remission state with a SLEDAI below 6 in the second sample collection point. With this information available, we investigated whether the prognosis of the patients could be predicted based on the methylation status of some CpG sites in samples from the first visit. We identified 494 DMPs in cMOs between patients who would remit and patients who would not. With these, we built a supervised learning k-nearest neighbours' algorithm to predict the progression of the disease. With an internal 10-fold cross-validation approach, we confirmed that 3 CpG positions were enough to build an accurate prediction algorithm (figure 3F,G).

Patients with SLE undergo drastic changes of monocyte subpopulations

To further investigate the presence of differences in our dataset, we leveraged single-cell RNA-seq dataset from PBMC samples of 162 patients with SLE and 98 HDs ([61]-GSE174188). Due to the difficulty of identifying iMOs in single-cell datasets, we initially divided the myeloid fraction of this dataset at a very low resolution, obtaining just two clusters: cluster 0 or CD16− cells and cluster 1 or CD16+ cells (figure 4C). With this initial division, DEGs from our bulk RNA-seq dataset significantly overlapped with the DEGs between SLE and HD cells in the single-cell object (cMO's DEGs in cluster 0, p=0.0096; iMO's and ncMO's DEGs in cluster 1, p=0.0164) (online supplemental figure 4A). Some of the genes upregulated in both datasets in the comparison of HD versus SLE included *IRF7*, *IFI27*, *HLA-A* or *HLA-C* (figure 4D and online supplemental figure 4B).

The density distribution of cells in the myeloid object exhibited a striking difference between SLE and HD samples. This contrast was especially pronounced in the SLE samples with higher activity (SLEDAI >6) (figure 4E). This finding prompted us to increase the resolution of the clustering in order to identify clusters containing cells from one specific condition (SLE or HD) (figure 4F). At a resolution of 0.15 (figure 4A), two of the new clusters mostly contained cells from the SLE samples. Specifically, clusters 1 and 6 contained 81% and 60% of cells from patients with SLE, respectively (figure 4A).

Analysis of some of the highly expressed genes for each cluster and their corresponding enriched pathways (figure 4F,G and online supplemental figure 4C) led us to the following annotations: cluster 0, primarily formed by cells of HD origin, was

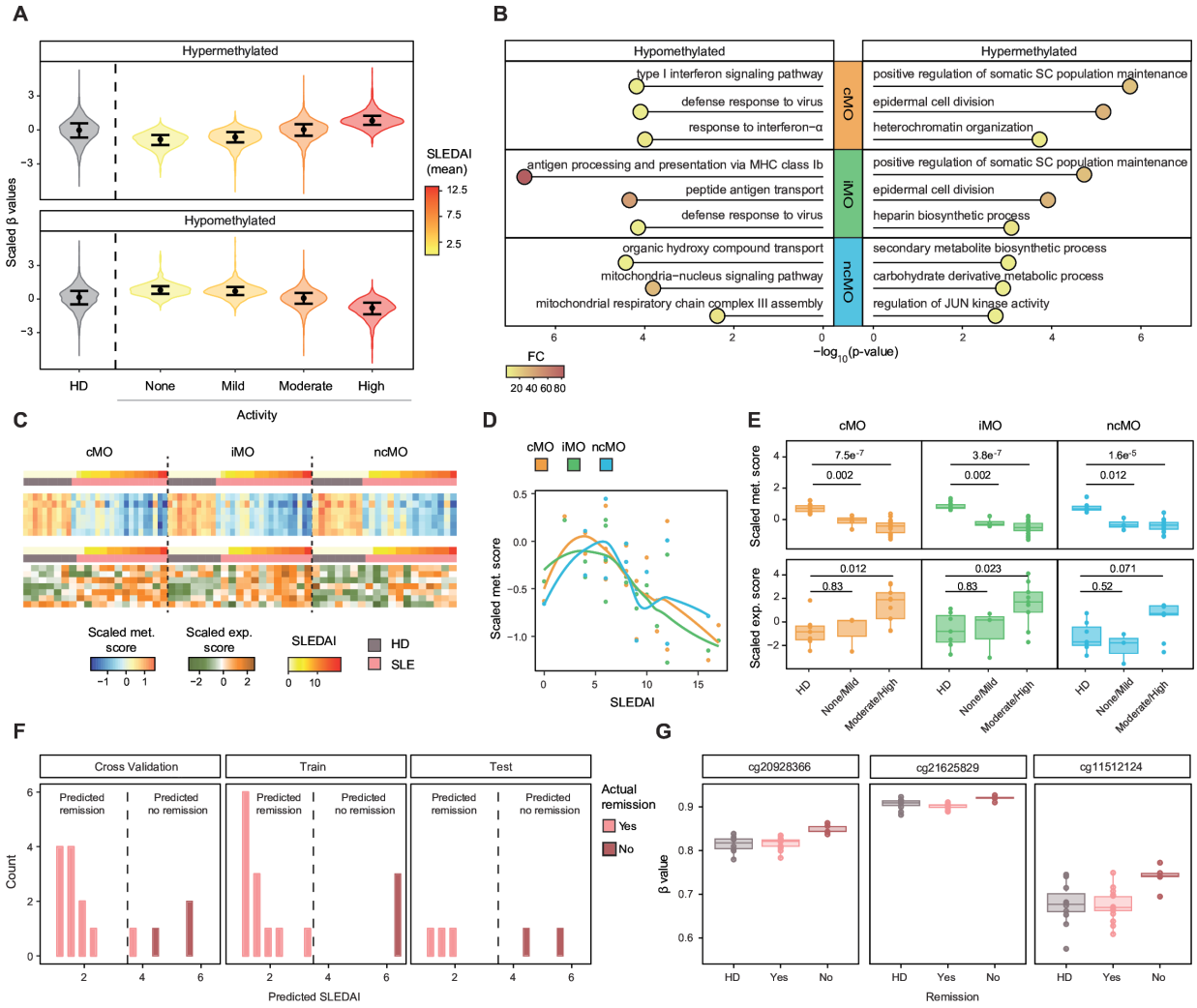


Figure 3 (A) Violin plot representation of the cMO's CpGs correlating with SLE activity as measured by SLEDAI in a Spearman's correlation. Samples divided in four groups according to their activity index: none is SLEDAI=0, mild is SLEDAI between 1 and 5, moderate is an index between 6 and 10, and high is a SLEDAI above 10. (B) Gene Ontology results of the genes annotated by proximity to the CpGs correlated with activity in each of the subsets. (C) Heatmap representation of STAT targets at the methylation (top) and expression (bottom) level. Methylation targets are defined per HOMER database and expression targets as per CollectRI database. Each row represents the results for a different STAT family member ordered from STAT1 in top row to STAT6 in bottom row. (D) Dotplot representation of methylation correlation with SLE activity index of the DMPs that are STAT1 targets in cluster M1. (E) Boxplot summary representation of STAT1 targets at the methylation (top) and expression (bottom) level in each of the subsets. Methylation targets are defined per HOMER database and expression targets as per CollectRI database. Samples grouped per activity so that none/mild group is SLEDAI 0–5 and moderate/high is SLEDAI above 6. Statistics performed with two-tailed Wilcoxon test. (F) Barplot representation of the results of the predictive model in the different cohorts. (G) DNA methylation levels of the three CpGs used by the predictive model in HD, remission patients and non-remission patients. cMO, classic monocyte; DMPs, differentially methylated positions; FC, fold change; HD, healthy donor; iMO, intermediate monocyte; MHC, major histocompatibility complex; ncMO, non-classic monocyte; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index.

labelled as S100A–mono and was enriched in pathways of response to LPS but also of response to oxidative stress. Cluster 1, essentially composed of SLE cells received the name of IFN–mono because it had an unmistakable IFN signature with *IFI6*, *IFI44L* and *ISG15* genes as its top markers. Cluster 2 was the main CD16+ cluster and, therefore, we labelled it as CD16+–mono. Cluster 2 was also enriched in pathways of response to type I IFN and regulation of leucocyte activation. Cluster 3 (MHC-II–mono) was strongly enriched for genes and pathways of the MHC-II and cluster 4 (EIF5A–mono) had higher levels of cell proliferation genes (*EFI4A* and *C1orf56*) and pathways related to cell adhesion and endocytosis. Cluster 5 (CCL3–mono) had a highly proinflammatory signature with marked TNF-α pathway activation. Finally, cluster 6 (C1Q–mono) highly expressed genes and pathways of the complement 1q (*C1QA, C1QB, C1QC*). Interestingly, MHC-II–mono, EIF5A–mono and C1Q–mono presented very similar features to those found in a previously annotated single-cell dataset of another autoimmune disease.[107]

By using CIBERSORTx software,[67] we studied the proportion of these clusters represented in our bulk dataset (figure 4H). With this approach, we determined that our bulk samples are composed of a heterogeneous mixture of cells showing differential patterns correlating with disease state and disease activity. As expected, cMOs from SLE donors were composed of a significant percentage of the labelled IFN–mono cluster, while HD samples barely harboured them. This representation was more elevated in samples with higher disease activity, where the

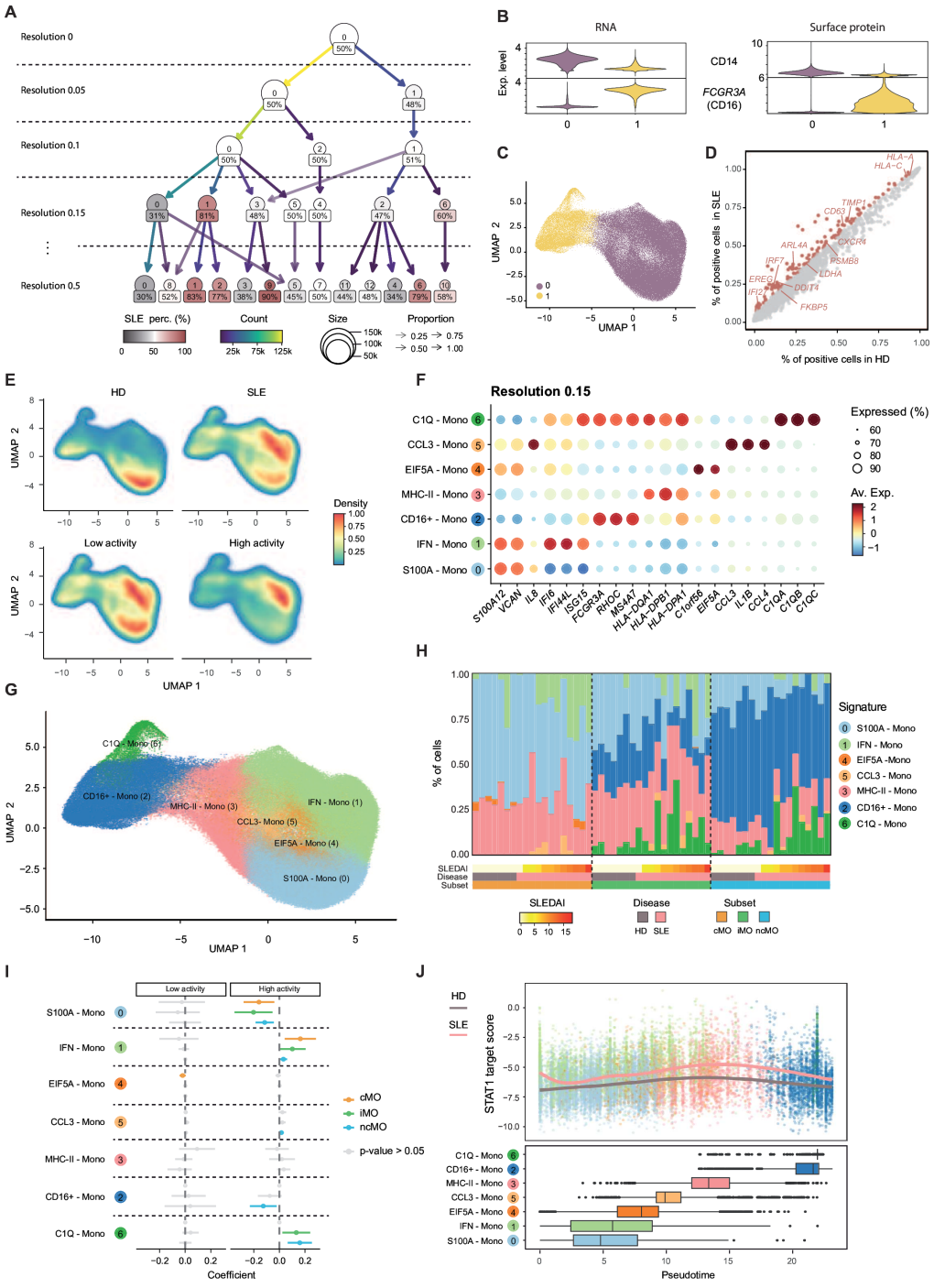


Figure 4 (A) Representation of unsupervised clustering results at different resolutions of the myeloid fraction of PBMCs from the single-cell RNA-seq dataset GSE174188. Node colour represents the percentage of SLE composition of each cluster. Edge colour represents the number of samples following each division. The transparency shows the incoming node proportion. (B) Violin plot representation of the expression levels of CD14 and CD16 (*FCGR3A*) at RNA and protein level in the two clusters generated by a 0.05 resolution. (C) UMAP representation of the distribution of the two clusters generated by a 0.05 resolution. (D) Scatter plot of the percentage of detected genes in HD and SLE samples in cluster 0 of the resolution 0.05. In colour are the DEGs from the single-cell RNA-seq dataset (FDR <0.05, $\log_2$FC >0.1). Labels present in DEGs commonly upregulated in the bulk dataset. (E) UMAP representation showing the density distribution of cells in the HD and SLE cohort (top), and within the SLE cohort, the low activity and high activity samples (bottom). SLEDAI of 6 was used as a threshold. (F) Dotplot representation of some of the DEGs in each cluster formed by a 0.15 resolution. (G) UMAP representation of the distribution of the six clusters generated by a 0.15 resolution. (H) Barplot representation of the estimate percentage composition of each bulk sample according to signature expression from each of the clusters generated by a 0.15 resolution. (I) Summary plot of percentage composition of bulk samples. Coefficients, p values and 95% CIs derived from fitting a linear model and testing it with Wilcoxon test. (J) Scatter plot representation of STAT1 targets expression score across the pseudotime distribution of the differentiation from CD16− to CD16+ monocytes (top). Summary representation of pseudotime representation of each cluster (bottom). cMO, classic monocyte; DEGs, differentially expressed genes; FDR, false discovery rate; HD, healthy donor; iMO, intermediate monocyte; $\log_2$FC, $\log_2$ fold change; ncMO, non-classic monocyte; PBMCs, peripheral blood mononuclear cells; RNA-seq, RNA sequencing; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index; UMAP, Uniform Manifold Approximation and Projection.

difference in proportions was statistically significant. We determined that iMOs contained a mixture of mostly S100A–mono, MHC-II–mono and CD16+–mono in HD samples, while iMO SLE samples repeatedly showed a non-trivial representation of C1Q–mono. This behaviour was similar also for ncMO samples, where C1Q–mono has a significantly higher proportion in the SLE samples of our bulk dataset. Another remarkable finding is the presence of the highly proinflammatory cluster CCL3–mono exclusively in SLE samples, being present in all three subsets from the bulk but never in HDs (figure 4I).

Finally, to inspect the impact of STAT1 dysregulation, initially determined from the bulk methylome and transcriptomic analyses, we studied the STAT1 regulon in the single-cell RNA-seq dataset. As anticipated, we observed an increased pattern of activation of this pathway in SLE samples. This activation was particularly exacerbated in the MHC-II–mono cluster and less in the S100A–mono or the CD16+ clusters (CD16+–mono and C1Q–mono) (figure 4J). Together, these results correlate with our previous findings of gene expression changes, indicating a higher STAT1 influence on the high disease activity samples, mainly in cMOs and iMOs, which are the samples with a heightened ratio of the MHC-II–mono cluster.

DISCUSSION

In this study, we have exploited the power of multiomics to dissect changes in monocyte subpopulations in SLE. Previous studies had suggested a differential role for CD16+ monocytes through variations in their proportions and functions compared with those in healthy individuals.[17–22] Our study sheds light on both the differential phenotype of monocyte subsets in SLE and the expansion of certain subpopulations. On one hand, DNA methylation profiles revealed changes between SLE and HD individuals that are both common to all monocyte subsets and unique to one of the subpopulations. On the other hand, transcriptomic analyses showed that cMOs were the subset with higher and more significant differences between SLE and HD. These differences, both at epigenetic and transcriptomic levels, correlated with disease activity, particularly the STAT family of TFs. We also developed a model to predict the prognosis of the disease by inspecting the DNA methylation values of only three CpG sites. Finally, by analysing a publicly available single-cell RNA-seq dataset of patients with SLE, we deconvoluted the composition of our bulk data and identified a signature of cells that explain the phenotype observed in the SLE monocyte subsets.

The involvement of monocyte subsets in SLE and other inflammatory conditions has been subject of debate in recent years. Discrepancies in results in monocyte proportions are evident: some studies indicate higher percentage of cMOs,[17] while others report an increase in iMOs[18 19] or in ncMOs.[20 21] Furthermore, certain studies have found no differences in the proportions of the subsets between SLE samples and HDs.[22] Our results indicate that the proportion of monocyte subsets varies in SLE in relation to disease activity. The proportion of iMOs is increased in SLE samples, particularly in those with low disease activity. The ncMO subset exhibits a similar trend, although statistical significance is not reached, likely due to cohort size limitations. These findings suggest a differential role of monocyte subsets at various stages of the disease.

DNA methylation changes can serve as indicators of environmental influence and can significantly affect cellular phenotype, predisposing cells to respond to subsequent stimuli. The chronic systemic inflammatory conditions characteristic of SLE contribute to shaping immune system cells into a state of constant responsiveness, perpetuating inflammation. Multiple studies have shown DNA methylation alterations in patients with SLE (reviewed in Ferreté-Bonastre *et al*[33]), generally pointing towards a dysregulated type I IFN signature. These modifications are associated with disease activity and prognosis. Our study on monocyte subsets has enabled us to identify epigenetic changes in positions shared among the different subsets, indicating a widespread inflammatory behaviour primarily driven by this cytokine. However, we also observed DMPs exclusive to each monocyte subpopulation. Notably, among the hypomethylated DMPs unique to cMOs, up to 80% harboured the consensus sequence for binding the Fos/Jun family of TFs. In the myeloid compartment, these TFs and the demethylation of their target sequences are associated with differentiation to macrophages.[108 109] This suggests a priming of these cells towards macrophage differentiation, which would occur instead of differentiation towards iMOs and ncMOs. In contrast, for DMPs exclusively demethylated in SLE ncMOs, we identified an association with T cell regulatory factors and pathways. Some of these DMPs were associated with the TF RORγt, a factor typically found in T cells, which has also been described in monocytes under some pathological conditions.[76] In this report, the expression of RORγt by a subset of monocytes was associated with their production of IL-17, a signature that resembles that of Th17 differentiation.[110] Interestingly, this type of T cell differentiation, particularly fostered by IFN-α-conditioned monocytes,[111] has been shown to play a role in SLE pathogenesis.[112 113] It is possible to speculate that the observed differences in monocyte subsets could be due to the differential response because of the cells expressing different receptors, of them being primed differently or that the bulk samples are composed of a heterogeneous mixture of different cells. It is important to note that although these analyses yielded biologically relevant results, they were conducted with an FDR of 0.2, potentially increasing the number of false positives. To mitigate this, we applied a beta difference threshold filter of 10%; however, this still represents a limitation of our study.

The transcriptomic analysis revealed that alterations were predominantly shared among the three subsets, while subset-specific changes were primarily observed in cMOs. This subset displayed more and greater differences in gene expression compared with the other subsets. This finding likely suggests that cMOs are in a less differentiated state, making them more plastic and better able to respond to an inflammatory environment. As anticipated, the genes upregulated in SLE were related to proinflammatory and antigen presentation pathways, whereas the downregulated genes were associated with negative regulation of autophagy or production of IL-12. These are relevant pathways associated with the development of SLE, where the dysregulated autophagosomic system may have an important role in the development and severity of the disease (reviewed in Liu *et al*[114]) and the IL-12/IL-23 axis is important for Th17 cell differentiation.[111] IFN signature, a well-known driver of SLE, was also present in our samples, particularly in those with higher activity as seen in both epigenetic and transcriptomic profile of STAT targets. In general terms, a pan-STAT signature was observed in samples with high activity, a finding supported by multiple reports suggesting the prominent role of this family of TFs in SLE (reviewed in Goropevšek *et al*[115]). In particular, at the transcriptomic level of STAT1 targets, the member of the family more closely related to IFN signalling, it appears that cMOs and iMOs are the most responsive and susceptible subsets. In contrast, ncMOs' STAT1 targets

expression does not reach a significant increase in high activity samples.

Based on the characteristics of our cohort and the clinical data collected regarding disease progression, we classified our patients into two groups: those with a good prognosis or in remission, and those with a poor prognosis or not in remission. Using this information, we developed a predictive model that used the epigenetic information from the first visit sample to predict the prognosis of the samples into these two categories. Remarkably, with the methylation status of just three CpG sites, we were able to achieve this classification in both our test and validation cohorts. This suggests that, even at the onset of a flare period, there are measurable indicators of the prognosis of the episode. A limitation of this aspect of our study is the size of our cohort, which necessitated division into independent groups for validation and testing. While these results should be further validated with a larger sample size, they are already promising, particularly considering the current sample size.

Finally, using a single-cell RNA-seq dataset from an SLE cohort, we identified subgroups of monocytes within our bulk dataset that help explain the differences observed. In this regard, cMOs and iMOs from SLE samples were formed by a high percentage of cells with an IFN signature, particularly in the high activity samples, which could explain the STAT1 signature that we had identified. Similarly, iMOs and ncMOs also contained a higher percentage of another group of cells with a high C1Q signature. This group of cells is likely to be important in SLE pathogenesis, where immunocomplexes formed by autoantibodies and complement factors are the main initiators of tissue damage. Of note, despite the results show that the bulk SLE samples are formed by a higher percentage of these proinflammatory groups of cells than the HD samples, they also contain an important representation of 'healthy' cells or cells present in HDs, it is not the whole of monocytes that are pathogenic. Hence, further investigation into the signals that influence a healthy monocyte, just released from the bone marrow, to develop a proinflammatory, disease-promoting cell phenotype, or to follow a physiological destiny, even in the inflammatory environment present in this disease, would be of great interest.

Author affiliations
[1]Epigenetics and Immune Disease Group, Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Spain
[2]Immunology Division, Vall d'Hebron University Hospital and Diagnostic Immunology Research Group, Vall d'Hebron Research Institute (VHIR), Barcelona, Spain
[3]Early Oncology Data Science, Computational Biology, AstraZeneca, Barcelona, Spain
[4]Cancer Epigenetics Group, Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Spain
[5]Centro de Investigación Biomédica en Red Cancer (CIBERONC), Madrid, Spain
[6]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
[7]Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Spain
[8]Rheumatology Department, Hospital Vall d'Hebron and Vall d'Hebron Research Institute (VHIR), Barcelona, Spain
[9]Epigenetics in Inflammatory and Metabolic Diseases Laboratory, Health Science Center (HSC), East China Normal University (ECNU), Shanghai, China

Acknowledgements We thank CERCA Programme/Generalitat de Catalunya and the Josep Carreras Foundation for institutional support. We also thank all the patients who graciously donated their time and samples for lupus research.

Contributors AGF-B and EB conceived experiments. AGF-B performed experiments. AGF-B, OM-P, CLC, JC-S and JR-U performed biocomputing analysis. AGF-B performed statistical analysis. MM-G and JC-H performed patient selection and provided samples. AGF-B, ME and EB analysed the data. AGF-B illustrated graphical representations. AGF-B and EB wrote the paper. EB is the guarantor of the study. All authors read and approved the final manuscript.

Funding This study was funded by the Spanish Ministry of Science and Innovation (PID2020-117212RB-I00/AEI/10.13038). EB was funded by the Spanish Ministry of Science and Innovation (MICINN; grant numbers PID2020-117212RB-I00/AEI/10.13038/501100011033). MM-G was funded by Project PI18/00346 (Instituto de Salud Carlos III and co-funded by European Union (ERDF/ESF).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Committee for Human Subjects of the Bellvitge Hospital Ethics Committee (PR275/17). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Methylation array and Expression data for this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO SuperSeries with accession number GSE249641.

ORCID iD
Esteban Ballestar http://orcid.org/0000-0002-1400-2440

REFERENCES

1 van Kempen TS, Wenink MH, Leijten EFA, *et al*. Perception of self: distinguishing autoimmunity from autoinflammation. *Nat Rev Rheumatol* 2015;11:483–92.
2 Ma W-T, Gao F, Gu K, *et al*. The role of monocytes and Macrophages in autoimmune diseases: a comprehensive review. *Front Immunol* 2019;10:1140.
3 Kwant LE, Vegting Y, Tsang-A-Sjoe MWP, *et al*. Macrophages in lupus nephritis: exploring a potential new therapeutic Avenue. *Autoimmun Rev* 2022;21:103211.
4 Rana AK, Li Y, Dang Q, *et al*. Monocytes in rheumatoid arthritis: circulating precursors of Macrophages and Osteoclasts and, their heterogeneity and plasticity role in RA pathogenesis. *Int Immunopharmacol* 2018;65:348–59.
5 Katsiari CG, Liossis S-NC, Souliotis VL, *et al*. Aberrant expression of the Costimulatory molecule Cd40 ligand on monocytes from patients with systemic lupus erythematosus. *Clin Immunol* 2002;103:54–62.
6 Harigai M, Hara M, Fukasawa C, *et al*. Responsiveness of peripheral blood B cells to recombinant Cd40 ligand in patients with systemic lupus erythematosus. *Lupus* 1999;8:227–33.
7 Decker P, Kötter I, Klein R, *et al*. Monocyte-derived Dendritic cells over-express Cd86 in patients with systemic lupus erythematosus. *Rheumatology (Oxford)* 2006;45:1087–95.
8 Herrmann M, Voll RE, Zoller OM, *et al*. Impaired Phagocytosis of apoptotic cell material by monocyte-derived Macrophages from patients with systemic lupus erythematosus. *Arthritis Rheum* 1998;41:1241–50.
9 Baumann I, Kolowos W, Voll RE, *et al*. Impaired uptake of apoptotic cells into Tingible body Macrophages in germinal centers of patients with systemic lupus erythematosus. *Arthritis Rheum* 2002;46:191–201.
10 Sampath P, Moideen K, Ranganathan UD, *et al*. Monocyte Subsets: phenotypes and function in tuberculosis infection. *Front Immunol* 2018;9:1726.
11 Kapellos TS, Bonaguro L, Gemünd I, *et al*. Human monocyte Subsets and phenotypes in major chronic inflammatory diseases. *Front Immunol* 2019;10:2035.
12 Patel AA, Zhang Y, Fullerton JN, *et al*. The fate and LifeSpan of human monocyte Subsets in steady state and systemic inflammation. *J Exp Med* 2017;214:1913–23.
13 Haschka D, Tymoszuk P, Bsteh G, *et al*. Expansion of neutrophils and classical and Nonclassical monocytes as a hallmark in relapsing-remitting multiple sclerosis. *Front Immunol* 2020;11:594.
14 Woollard KJ, Geissmann F. Monocytes in Atherosclerosis: Subsets and functions. *Nat Rev Cardiol* 2010;7:77–86.
15 Godoy-Tena G, Barmada A, Morante-Palacios O, *et al*. Epigenetic and Transcriptomic Reprogramming in monocytes of severe COVID-19 patients reflects alterations in myeloid differentiation and the influence of inflammatory Cytokines. *Genome Med* 2022;14:134.
16 Silvin A, Chapuis N, Dunsmore G, *et al*. Elevated Calprotectin and abnormal myeloid cell Subsets discriminate severe from mild COVID-19. *Cell* 2020;182:S0092-8674(20)30993-4:1401–1418..
17 Wu Z, Zhang S, Zhao L, *et al*. Upregulation of Cd16- monocyte Subsets in systemic lupus erythematous patients. *Clin Rheumatol* 2017;36:2281–7.

18 Zhu H, Hu F, Sun X, *et al*. Cd16 + monocyte subset was enriched and functionally exacerbated in driving T-cell activation and B-cell response in systemic lupus erythematosus. *Front Immunol* 2016;7:512.
19 López P, Rodríguez-Carrio J, Martínez-Zapico A, *et al*. Low-density Granulocytes and monocytes as biomarkers of cardiovascular risk in systemic lupus erythematosus. *Rheumatology (Oxford)* 2020;59:1752–64.
20 Mukherjee R, Kanti Barman P, Kumar Thatoi P, *et al*. Non-classical monocytes display inflammatory features: validation in sepsis and systemic lupus erythematous. *Sci Rep* 2015;5:13886.
21 Jiang W, Zhang L, Lang R, *et al*. Sex differences in monocyte activation in systemic lupus erythematosus (SLE). *PLoS ONE* 2014;9:e114589.
22 Li Y, Lee PY, Sobel ES, *et al*. Increased expression of Fcgammari/Cd64 on circulating monocytes parallels ongoing inflammation and nephritis in lupus. *Arthritis Res Ther* 2009;11:R6.
23 Skrzeczyńska-Moncznik J, Bzowska M, Loseke S, *et al*. Peripheral blood Cd14High Cd16+ monocytes are main producers of IL-10. *Scand J Immunol* 2008;67:152–9.
24 Szaflarska A, Baj-Krzyworzeka M, Siedlar M, *et al*. Antitumor response of Cd14+/ Cd16+ monocyte subpopulation. *Exp Hematol* 2004;32:748–55.
25 Belge K-U, Dayyani F, Horelt A, *et al*. The proinflammatory Cd14+Cd16+DR++ monocytes are a major source of TNF. *J Immunol* 2002;168:3536–42.
26 Boyette LB, Macedo C, Hadi K, *et al*. Phenotype, function, and differentiation potential of human monocyte Subsets. *PLoS One* 2017;12:e0176460.
27 Narasimhan PB, Marcovecchio P, Hamers AAJ, *et al*. Nonclassical monocytes in health and disease. *Annu Rev Immunol* 2019;37:439–56.
28 Auffray C, Fogg D, Garfa M, *et al*. Monitoring of blood vessels and tissues by a population of monocytes with patrolling behavior. *Science* 2007;317:666–70.
29 Yona S, Kim K-W, Wolf Y, *et al*. Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* 2013;38:79–91.
30 Byrne JC, Ní Gabhann J, Lazzari E, *et al*. Genetics of SLE: functional relevance for monocytes/Macrophages in disease. *Clin Dev Immunol* 2012;2012:582352.
31 Charras A, Smith E, Hedrich CM. Systemic lupus erythematosus in children and young people. *Curr Rheumatol Rep* 2021;23:20.
32 Barbhaiya M, Costenbader KH. Environmental exposures and the development of systemic lupus erythematosus. *Curr Opin Rheumatol* 2016;28:497–505.
33 Ferreté-Bonastre AG, Cortés-Hernández J, Ballestar E. What can we learn from DNA methylation studies in lupus. *Clin Immunol* 2022;234:108920.
34 Postal M, Vivaldo JF, Fernandez-Ruiz R, *et al*. Type I interferon in the pathogenesis of systemic lupus erythematosus. *Curr Opin Immunol* 2020;67:87–94.
35 Ballestar E, Sawalha AH, Lu Q. Clinical value of DNA methylation markers in autoimmune rheumatic diseases. *Nat Rev Rheumatol* 2020;16:514–24.
36 Aringer M, Costenbader K, Daikh D, *et al*. European League against rheumatism/ American college of rheumatology classification criteria for systemic lupus erythematosus. *Arthritis Rheumatol* 2019;71:1400–12.
37 Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol* 2002;29:288–91.
38 Petri M, Buyon J, Kim M. Classification and definition of major flares in SLE clinical trials. *Lupus* 1999;8:685–91.
39 Buyon JP, Petri MA, Kim MY, *et al*. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. *Ann Intern Med* 2005;142(12 Pt 1):953–62.
40 Ziegler-Heitbrock L, Ancuta P, Crowe S, *et al*. Nomenclature of monocytes and Dendritic cells in blood. *Blood* 2010;116:e74–80.
41 de la Calle-Fabregat C, Rodríguez-Ubreva J, Cañete JD, *et al*. Designing studies for epigenetic biomarker development in autoimmune rheumatic diseases. *Rheumatology and Immunology Research* 2022;3:103–10.
42 Andrews S. Fastqc: A quality control tool for high throughput sequence data; 2010.
43 Morante-Palacios O, Ballestar E. Shinyépico: a graphical pipeline to analyze Illumina DNA methylation arrays. *Bioinformatics* 2021;37:257–9.
44 Du P, Kibbe WA, Lin SM. Lumi: a pipeline for processing Illumina Microarray. *Bioinformatics* 2008;24:1547–8.
45 Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina Humanmethylationepic array with Minfi. *Bioinformatics* 2017;33:558–60.
46 Ritchie ME, Phipson B, Wu D, *et al*. Limma powers differential expression analyses for RNA-sequencing and Microarray studies. *Nucleic Acids Res* 2015;43:e47.
47 Yu G, Wang LG, He QY. Chipseeker: an R/Bioconductor package for chip peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–3.
48 Wang Q, Li M, Wu T, *et al*. Exploring Epigenomic Datasets by Chipseeker. *Curr Protoc* 2022;2:e585.
49 Heinz S, Benner C, Spann N, *et al*. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89.
50 Gu Z, Hübschmann D. rGREAT: an R/Bioconductor package for functional enrichment on Genomic regions. *Bioinformatics* 2023;39:btac745.
51 Cavalcante RG, Sartor MA. Annotatr: Genomic regions in context. *Bioinformatics* 2017;33:2381–3.
52 Morgan M, AnnotationHub SL. Client to access Annotationhub resources; *R package version 3.4 .2022*,
53 Bray NL, Pimentel H, Melsted P, *et al*. Near-optimal probabilistic RNA-Seq Quantification. *Nat Biotechnol* 2016;34:525–7.
54 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with Deseq2. *Genome Biol* 2014;15:550.
55 Wickham H. Ggplot2. In: *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer-Verlag New York, 2016.
56 Yu G, Wang L-G, Han Y, *et al*. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
57 Carlson M. Org.Hs.Eg.dB: genome wide annotation for human; *R Package Version 382 2019*,
58 Alvarez MJ, Shen Y, Giorgi FM, *et al*. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* 2016;48:838–47.
59 Müller-Dott S, Tsirvouli E, Vazquez M, *et al*. Expanding the coverage of Regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Res* 2023;51:10934–49.
60 Garcia-Alonso L, Holland CH, Ibrahim MM, *et al*. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* 2019;29:1363–75.
61 Perez RK, Gordon MG, Subramaniam M, *et al*. Single-cell RNA-Seq reveals cell type-specific molecular and genetic associations to lupus. *Science* 2022;376:eabf1970.
62 Hao Y, Hao S, Andersen-Nissen E, *et al*. Integrated analysis of Multimodal single-cell data. *Cell* 2021;184:S0092-8674(21)00583-3:3573–3587..
63 Zappia L, Oshlack A. Clustering trees: a visualization for evaluating Clusterings at multiple resolutions. *Gigascience* 2018;7:giy083.
64 Trapnell C, Cacchiarelli D, Grimsby J, *et al*. The Dynamics and regulators of cell fate decisions are revealed by Pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381–6.
65 Qiu X, Mao Q, Tang Y, *et al*. Reversed graph Embedding resolves complex single-cell Trajectories. *Nat Methods* 2017;14:979–82.
66 Cao J, Spielmann M, Qiu X, *et al*. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 2019;566:496–502.
67 Newman AM, Steen CB, Liu CL, *et al*. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37:773–82.
68 Oikawa D, Gi M, Kosako H, *et al*. Otud1 Deubiquitinase regulates NF-κB- and Keap1-mediated inflammatory responses and reactive oxygen species-associated cell death pathways. *Cell Death Dis* 2022;13:694.
69 Rochford I, Joshi JC, Rayees S, *et al*. Evidence for Reprogramming of monocytes into Reparative alveolar Macrophages in vivo by targeting Pde4B. *Am J Physiol Lung Cell Mol Physiol* 2021;321:L686–702.
70 Kramer PR, Winger V, Reuben J. Pi3K limits TNF-alpha production in Cd16-activated monocytes. *Eur J Immunol* 2009;39:561–70.
71 Guha M, Mackman N. The Phosphatidylinositol 3-kinase-AKT pathway limits Lipopolysaccharide activation of signaling pathways and expression of inflammatory mediators in human Monocytic cells. *J Biol Chem* 2002;277:32124–32.
72 Gerszten RE, Friedrich EB, Matsui T, *et al*. Role of Phosphoinositide 3-kinase in monocyte recruitment under flow conditions. *J Biol Chem* 2001;276:26846–51.
73 Nakamura T, Datta R, Kharbanda S, *et al*. Regulation of Jun and Fos gene expression in human monocytes by the macrophage colony-stimulating factor. *Cell Growth Differ* 1991;2:267–72.
74 Zenz R, Eferl R, Scheinecker C, *et al*. Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease. *Arthritis Res Ther* 2008;10:201.
75 Castro G, Liu X, Ngo K, *et al*. RORγt and RORα signature genes in human Th17 cells. *PLoS One* 2017;12:e0181868.
76 Lourda M, Olsson-Åkefeldt S, Gavhed D, *et al*. Detection of IL-17A-producing peripheral blood monocytes in Langerhans cell Histiocytosis patients. *Clin Immunol* 2014;153:112–22.
77 Kim HA, Lee HS, Shin TH, *et al*. Polyamine patterns in plasma of patients with systemic lupus erythematosus and fever. *Lupus* 2018;27:930–8.
78 Dos Santos M, Veronese FV, Moresco RN. Uric acid and kidney damage in systemic lupus erythematosus. *Clin Chim Acta* 2020;508:197–205.
79 Coit P, Yalavarthi S, Ognenovski M, *et al*. Epigenome profiling reveals significant DNA Demethylation of interferon signature genes in lupus neutrophils. *J Autoimmun* 2015;58:59–66.
80 Absher DM, Li X, Waite LL, *et al*. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent Hypomethylation of interferon genes and compositional changes to Cd4+ T-cell populations. *PLoS Genet* 2013;9:e1003678.
81 Miller S, Tsou PS, Coit P, *et al*. Hypomethylation of Stat1 and HLA-Drb1 is associated with type-I interferon-dependent HLA-Drb1 expression in lupus Cd8+ T cells. *Ann Rheum Dis* 2019;78:519–28.
82 Coit P, Jeffries M, Altorok N, *et al*. Genome-wide DNA methylation study suggests epigenetic accessibility Andtranscriptional poising of interferon-regulated genes in Naïve Cd4+ T Cellsfrom lupus patients. *J Autoimmun* 2013;43:78–84.
83 Sawalha AH, Wang L, Nadig A, *et al*. Sex-specific differences in the relationship between genetic susceptibility, T cell DNA Demethylation and lupus flare severity. *J Autoimmun* 2012;38:J216–22.

84 Zhang Y, Liu Q, Yang S, *et al*. Cd58 Immunobiology at a glance. *Front Immunol* 2021;12.
85 Hoffmann JC, Räuker HJ, Krüger H, *et al*. Decreased levels of a soluble form of the human adhesion receptor Cd58 (LFA-3) in sera and Synovial fluids of patients with rheumatoid arthritis. *Clin Exp Rheumatol* 1996;14:23–9.
86 De Jager PL, Baecher-Allan C, Maier LM, *et al*. The role of the Cd58 locus in multiple sclerosis. *Proc Natl Acad Sci U S A* 2009;106:5264–9.
87 Mende R, Vincent FB, Kandane-Rathnayake R, *et al*. Analysis of serum interleukin (IL)-1**B** and IL-18 in systemic lupus erythematosus. *Front Immunol* 2018;9:1250.
88 Gutzmer R, Langer K, Mommert S, *et al*. Human Dendritic cells express the IL-18R and are Chemoattracted to IL-18. *J Immunol* 2003;171:6363–71.
89 Gauzzi MC, Purificato C, Conti L, *et al*. IRF-4 expression in the human myeloid lineage: up-regulation during Dendritic cell differentiation and inhibition by 1Alpha,25-dihydroxyvitamin D3. *J Leukoc Biol* 2005;77:944–7.
90 Jin S, Song Y, Zhou L, *et al*. Depletion of Cul4B in Macrophages ameliorates diabetic kidney disease via miR-194-5p/Itga9 axis. *Cell Reports* 2023;42:112550.
91 Van Eyck L, De Somer L, Pombal D, *et al*. Brief report: Ifih1 mutation causes systemic lupus erythematosus with selective IgA deficiency. *Arthritis Rheumatol* 2015;67:1592–7.
92 Hess NJ, Felicelli C, Grage J, *et al*. Tlr10 suppresses the activation and differentiation of monocytes with effects on DC-mediated adaptive immune responses. *J Leukoc Biol* 2017;101:1245–52.
93 Huang X, Hua J, Shen N, *et al*. Dysregulated expression of Interleukin-23 and Interleukin-12 subunits in systemic lupus erythematosus patients. *Mod Rheumatol* 2007;17:220–3.
94 Qiu F, Song L, Yang N, *et al*. Glucocorticoid Downregulates expression of IL-12 family Cytokines in systemic lupus erythematosus patients. *Lupus* 2013;22:1011–6.
95 Larosa M, Zen M, Gatto M, *et al*. IL-12 and IL-23/Th17 axis in systemic lupus erythematosus. *Exp Biol Med (Maywood)* 2019;244:42–51.
96 Tang Y-Y, Wang D-C, Wang Y-Q, *et al*. Emerging role of hypoxia-inducible Factor-1**A** in inflammatory autoimmune diseases: a comprehensive review. *Front Immunol* 2022;13:1073971.
97 Álvarez K, Villar-Vesga J, Ortiz-Reyes B, *et al*. Induction of NF-**K**B inflammatory pathway in monocytes by microparticles from patients with systemic lupus erythematosus. *Heliyon* 2020;6:e05815.
98 Brightbill HD, Suto E, Blaquiere N, *et al*. NF-**K**B inducing kinase is a therapeutic target for systemic lupus erythematosus. *Nat Commun* 2018;9:179.
99 Barnabei L, Laplantine E, Mbongo W, *et al*. NF-**K**B: at the borders of autoimmunity and inflammation. *Front Immunol* 2021;12:716469.
100 Trizzino M, Zucco A, Deliard S, *et al*. Egr1 is a Gatekeeper of inflammatory enhancers in human macrophages. *Sci Adv* 2021;7:eaaz8836.
101 de la Calle-Fabregat C, Rodríguez-Ubreva J, Ciudad L, *et al*. The Synovial and blood monocyte DNA Methylomes mirror prognosis, evolution, and treatment in early arthritis. *JCI Insight* 2022;7:e158783.
102 de la Calle-Fabregat C, Niemantsverdriet E, Cañete JD, *et al*. Prediction of the progression of undifferentiated arthritis to rheumatoid arthritis using DNA methylation profiling. *Arthritis Rheumatol* 2021;73:2229–39.
103 Mok A, Rhead B, Holingue C, *et al*. Hypomethylation of Cyp2E1 and Dusp22 promoters associated with disease activity and erosive disease among rheumatoid arthritis patients. *Arthritis Rheumatol* 2018;70:528–36.
104 Zhao M, Zhou Y, Zhu B, *et al*. Ifi44L promoter methylation as a blood biomarker for systemic lupus erythematosus. *Ann Rheum Dis* 2016;75:1998–2006.
105 Lu M-C, Lai N-S, Chen H-C, *et al*. Decreased microRNA(miR)-145 and increased miR-224 expression in T cells from patients with systemic lupus erythematosus involved in lupus Immunopathogenesis. *Clin Exp Immunol* 2013;171:91–9.
106 Karonitsch T, Feierl E, Steiner CW, *et al*. Activation of the interferon-gamma signaling pathway in systemic lupus erythematosus peripheral blood mononuclear cells. *Arthritis Rheum* 2009;60:1463–71.
107 Zheng W, Wang X, Liu J, *et al*. Single-cell analyses highlight the proinflammatory contribution of C1Q-high monocytes to Behçet's disease. *Proc Natl Acad Sci U S A* 2022;119:e2204289119.
108 Dekkers KF, Neele AE, Jukema JW, *et al*. Human monocyte-to-macrophage differentiation involves highly localized gain and loss of DNA methylation at transcription factor binding sites. *Epigenetics Chromatin* 2019;12:34.
109 Zhao X, Bartholdy B, Yamamoto Y, *et al*. PU.1-C-Jun interaction is crucial for PU.1 function in myeloid development. *Commun Biol* 2022;5.
110 Chong WP, Mattapallil MJ, Raychaudhuri K, *et al*. The cytokine IL-17A limits Th17 Pathogenicity via a negative feedback loop driven by Autocrine induction of IL-24. *Immunity* 2020;53:S1074-7613(20)30277-6:384–397..
111 Santini SM, Lapenta C, Donati S, *et al*. Interferon-**A**-conditioned human monocytes combine a Th1-orienting attitude with the induction of autologous Th17 responses: role of IL-23 and IL-12. *PLoS One* 2011;6:e17364.
112 Talaat RM, Mohamed SF, Bassyouni IH, *et al*. Th1/Th2/Th17/Treg cytokine imbalance in systemic lupus erythematosus (SLE) patients: correlation with disease activity. *Cytokine* 2015;72:146–53.
113 Abou Ghanima AT, Elolemy GG, Ganeb SS, *et al*. Role of T helper 17 cells in the pathogenesis of systemic lupus erythematosus. *Egypt J Immunol* 2012;19:25–33.
114 Liu X, Qin H, Xu J. The role of Autophagy in the pathogenesis of systemic lupus erythematosus. *Int Immunopharmacol* 2016;40:351–61.
115 Goropevšek A, Holcar M, Avčin T. The role of STAT signaling pathways in the pathogenesis of systemic lupus erythematosus. *Clin Rev Allergy Immunol* 2017;52:164–81.

Clinical Immunology 234 (2022) 108920

Contents lists available at ScienceDirect

Clinical Immunology

journal homepage: www.elsevier.com/locate/yclim

Review Article

What can we learn from DNA methylation studies in lupus?

Anna Guiomar Ferreté-Bonastre [a], Josefina Cortés-Hernández [b], Esteban Ballestar [a,c,*]

[a] Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain
[b] Department of Rheumatology, Vall d'Hebron Research Institute (VHIR), 08035, Barcelona, Spain
[c] Epigenetics in Inflammatory and Metabolic Diseases Laboratory, Health Science Center (HSC), East China Normal University (ECNU), Shanghai, 200241, China

ARTICLE INFO

Keywords:
DNA methylation
Lupus
Epigenetics
Clinical applications

ABSTRACT

During the past twenty years, a wide range of studies have established the existence of epigenetic alterations, particularly DNA methylation changes, in lupus. Epigenetic changes might have different contributions in children-onset versus adult-onset lupus. DNA methylation alterations have been identified and characterized in relation to disease activity and damage, different lupus subtypes and responses to drugs. However, to date there has been no practical application of these findings in the clinical milieu. In this article, we provide a review of key studies showing the relationship between DNA methylation and the many clinical aspects related to lupus. We also propose several options, in relation to the range of methodological developments and experimental design, that could optimize these findings and make them amenable for use in clinical practice.

1. Introduction

Lupus is an archetypal systemic autoimmune disease characterized by the production of autoantibodies, complement activation, and immune complex formation and deposition [1]. It has an exceptionally heterogeneous presentation among affected individuals, making it challenging sometimes for clinicians to diagnose. With a prevalence ranging from 20 to 200 cases per 100,000 people in distinct ancestral populations [2–5], many efforts have been made to better understand the pathogenesis and development of this complex disease.

The etiology of lupus is believed to be a combination of genetic and environmental factors. Genetic contributions are thought to be particularly determinant in childhood-onset lupus, which represent 10–20% of lupus patients (reviewed in [6]) since these patients have been exposed to fewer, if any, triggering environmental impacts. Over the last two decades, genome-wide association studies (GWASs) have identified an association between lupus and certain alleles at more than 130 different loci [7–9], suggesting a multiple-gene susceptibility to disease development. Many of the susceptibility genes, such as *IFIH1* [8,10], and *IL10* [11], are directly linked to immune system functions. Some, like *PTPN22* [12,13], CD80 [14], and *IL12A* [8], are characteristic of T cell functions; others, like *BANK1* [9,15], *PRDM1* [11], and *BLK* [15,16], are typical of B cell functions; yet others, such as *ITGAM* [15,17] and *ICAMs* [18], are typically expressed in the myeloid lineage. This variety of immune functions associated with genetic susceptibility underlines the importance of different cell types in lupus pathogenesis. Childhood-onset lupus has been associated in multiple studies with higher presence of SLE-associated polymorphism and higher genetic risk [19–21] Moreover, 1–4% of total SLE patients and 7–8% in childhood-onset SLE [22] present a *monogenic* form of lupus, with high penetrance single mutations that cause patients to develop the disease. These rare mutations take part in complement components, endonucleases genes or in apoptotic and lymphocyte activation pathways [23–26].

There are also several lines of evidence of the activity of non-genetic factors in lupus development (reviewed in [27]). For instance, the frequent occurrence of discordance for lupus in monozygotic twins is evidence of the presentation of non-genetic factors [28–30]. In this regard, environmental factors such as cigarette smoking [31,32], hormones [33,34] or infections [35], among others, are thought to be correlated with disease development (reviewed in [36]). Currently, lupus is considered a complex disease associated with multigenetic susceptibility and environmental triggers. In general, genetic factors seem to have a higher impact in childhood-onset lupus patients while adult-onset patients are thought to involve more environmental factors in order to develop the disease (reviewed in [37]).

A possible mechanism of action of environmental triggers on the phenotype is through epigenetics. Epigenetics provides several mechanisms that can determine variability in gene expression and, ultimately, function without involving changes in the genetic sequence. In this regard, methylation of the 5′ position of the pyrimidine ring of cytosine

* Corresponding author at: Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain.
E-mail address: eballestar@carrerasresearch.org (E. Ballestar).

https://doi.org/10.1016/j.clim.2021.108920
Received 3 November 2021; Received in revised form 23 December 2021; Accepted 26 December 2021
Available online 29 December 2021

residues adjacent to guanines (CpG dinucleotides) (from now on DNA methylation) and the post-translational modification of amino acid residues in histones are directly associated with gene expression. The covalent modification of the DNA and the histones that pack it are targeted to specific gene sequences by transcription factors and linked to upstream signaling pathways that depend, in some cases, on extracellular traits. This gives rise to direct relationships between environmental cues and gene-specific alterations of gene expression [38]. Epigenetic regulation is tightly regulated and takes different forms depending on the genomic region studied (see representative examples in Fig. 1A). DNA methylation is the most extensively studied category of epigenetic modification, and is particularly convenient for clinical studies because of the stability of the DNA molecule.

The processes of DNA methylation incorporation and removal involve a variety of enzymes and a series of complex interactions between enzymes such as DNA methyltransferases (DNMTs) and ten-eleven translocation enzymes (TETs), and transcription factors (Fig. 1B). These mechanisms are essential for defining function and can be altered under pathological conditions [39] (Fig. 2). In cancer, DNA methylation alterations have been extensively studied and are an active area of research because of their important roles in pathogenesis and the generation of clinically valuable markers (reviewed in [40]). In lupus research, pioneer studies in the 1990s established the existence of DNA methylation alterations in T cells [41]. In relation to monozygotic twin discordance, our team demonstrated the widespread occurrence of hypomethylation [42], and was the first to use high-throughput approaches to identify DNA methylation alterations in autoimmune disease. Apart from being altered under pathological conditions, this epigenetic modification is also drastically modified by age in all tissues. Particularly in the immune system, DNA methylation-related changes can lead to age-related dysfunctions (reviewed in [43]). In fact, some studies have pointed towards an age-associated DNA methylation dysregulation that could eventually predispose for autoimmunity [44,45].

During the past decade, there have been many studies of DNA methylation changes in lupus in relation to various clinical aspects, including disease susceptibility and activity, lupus subtypes, and response to drugs. Despite the evidence of their value, these markers have yet to be implemented in clinical practice. In the following sections, we consider representative studies focusing on DNA methylation and lupus that highlight the aspects of lupus pathogenesis and treatment that are significant from an epigenetic standpoint (summarized in Table 1). Finally, we provide a perspective on some considerations that may improve our knowledge so that it can be applied in clinical practice.

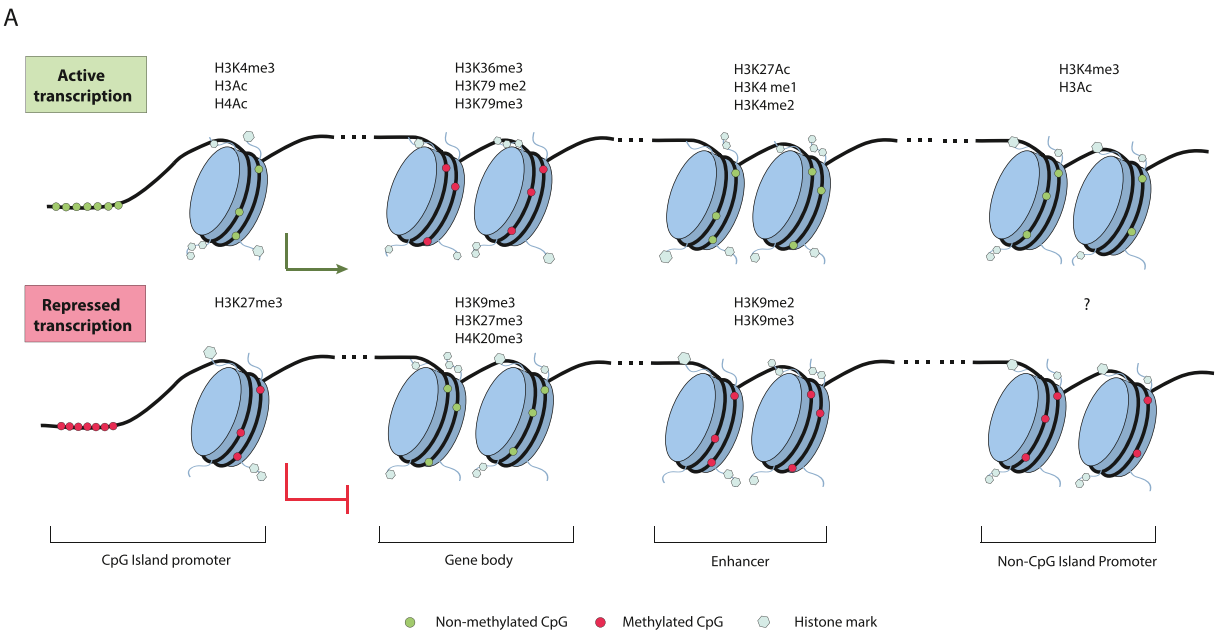
2. DNA methylation and disease susceptibility

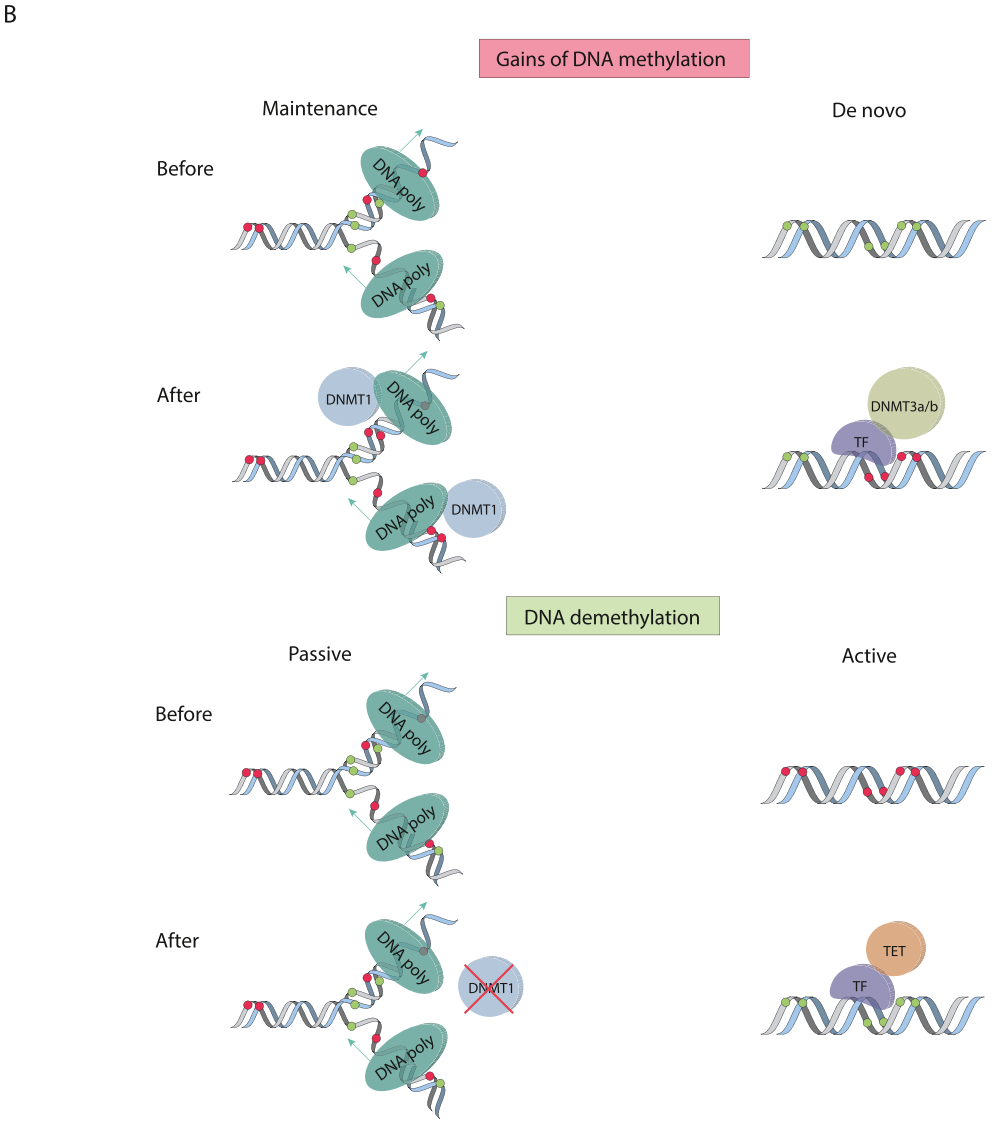
Similar to the aforementioned studies that addressed the genetic predisposition to developing SLE arising from certain combinations of single-nucleotide polymorphisms (SNPs) in various genes, other researchers have described an epigenetic susceptibility to lupus, with most of their studies largely focusing on DNA methylation. These results can be grouped with respect to three predisposing principles: the potential occurrence of epigenetically poised progenitor cells; the occurrence of alterations in the X-chromosome; and abnormal DNA methylation associated with SNP variation.

It is important to note that most epigenome-wide studies in the context of DNA methylation in SLE have predominantly described hypomethylated profiles [41,46]. Hypomethylation takes place especially in genes of the interferon-signaling pathway. Interferon has historically been demonstrated to be involved in lupus pathogenesis (reviewed in [47]). An interferon signature expression profile has been known to be associated with lupus patients for decades [48]. Moreover, monogenic forms of lupus have also been linked to interferon pathways. For example, rare homozygous mutations in *C1q* gene cause lupus with a high penetrance. This complement component is responsible of inhibiting IFNα production by innate immune cells [49]. Not only monogenic forms, but also susceptibility alleles have also been described in Type I interferon pathway (reviewed in [47]), linking once again this pathway with the development of the disease. In the case of DNA methylation, the hypomethylated profile on interferon-signaling pathway is also relevant in neutrophils and low-density granulocytes [50], CD19+ B cells and CD14+ monocytes [51], CD8+ T cells [52] and memory, regulatory and naïve CD4+ T cells [51,53,54]. The authors of several of these studies argue that the occurrence of this hypomethylation profile in interferon genes in all the cell types studied might be explained by the existence of a multipotent progenitor that bears this epigenetic alteration [51,53]. The response of susceptible individuals carrying these epigenetically poised progenitor cells would be magnified, as well as give rise to descendant cells that respond strongly to interferon.

As is the case for most autoimmune diseases, lupus occurs more frequently in females than in males, with ratios around 9:1 for adult-onset patients [55] and 6:1 for childhood-onset patients [56]. Many factors are believed to be responsible for this prominent imbalance, such as sex hormones, environmental factors, and reproductive history [57]. X chromosome dosage has also been proposed to be a key element predisposing for this disease, as demonstrated by the varying prevalence rates among abnormal karyotype individuals (47 XXY males [58], 47 XXX females [59], and 45 XO females [60]). In fact, men are reported to require a higher genetic risk/DNA methylation ratio to be as susceptible as women to developing SLE [54]. The X chromosome is enriched in genes involved in immune pathways [61], so it is not surprising that dysregulation of X chromosome inactivation mechanisms has severe consequences in the form of immune system-related susceptibilities. This is the case for genes like *OGT*, *CXCR3* [62], and *CD40LG* [63]. Richardson's group described all three X-linked genes as being regulated by DNA methylation. CD4+ T cells from women with lupus presented significantly greater hypomethylation and overexpression than those of men with lupus. Remarkably, all three mentioned genes could have a role in immune cells regulation. *CD40LG* has been clearly described to have a critical role in the pathogenesis of SLE due to the overstimulation of autologous B cells that produce IgG [64]. CXCR3 is a chemokine receptor found in Th1 CD4+ effector T cells that has been proved to be relevant in lupus pathogenesis, especially in lupus nephritis development, by being a firm candidate of CD4+ T cell recruitment to inflamed kidney tissue [65]. OGT, on the other hand, is an *N*-acetylglucosamine transferase, an epigenetic regulator [66], that is required for T and B cell activation [67]. It is reasonable, thus, to suggest that the aberrant DNA methylation of these genes may lead to a dysregulation of immune system pathways through the abnormal regulation of gene expression, contributing to the pathogenesis of SLE.

Finally, the development of GWAS has led to the emergence of a new type of analysis of DNA methylation patterns. Several studies have been published that investigate the genetic basis of DNA methylation variations and the possible role they may have in disease susceptibility. In fact, around 20% of the inter-individual variability in DNA methylation may be explained by genetics [68,69]. This layer of variance between individuals could account for the different incidences of SLE across ancestry groups [70,71]. For example, *cis*-methylation quantitative trait loci (cis-meQTL) analysis revealed 466 positions that were differentially methylated between SLE patients and controls that were genetically regulated by a SNP located within a distance of 1 Mb. Some of the associations were annotated with previously reported susceptibility-linked loci such as *IRF5*, *IRF7*, and *UBE2L3* [72]. The same was true for the *BANK1* gene, in which the methylation level of the CpG cg01116491 in this locus was correlated with several SNPs. Patients with the risk allele in homozygosis presented hypermethylation of this CpG relative to patients who were heterozygous or homozygous for the non-risk allele [73]. These findings shed light on the complex topic of genetic susceptibility to SLE and suggest potential mechanisms by which this variability translates into phenotypic differences.





(caption on next page)

Fig. 1. A. Schematic representation showing the chromatin and the profiles of histone post-translation modifications and DNA methylation in different genomic regions, and their relationship with gene transcription. Non-methylated CpGs are represented as green circles, methylated CpGs as red circles. The different histone marks are collectively represented by blue hexagons. B. Schematic representation of the mechanisms that lead to methylated (upper half) or non-methylated (lower half) DNA. Non-methylated CpGs in green circles, methylated CpGs in red circles. DNA poly, DNA polymerase; DNMT1 and 3a/b, DNA methyltransferases 1 and 3a or 3b, respectively; TET, ten-eleven translocation enzyme; TF, transcription factor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

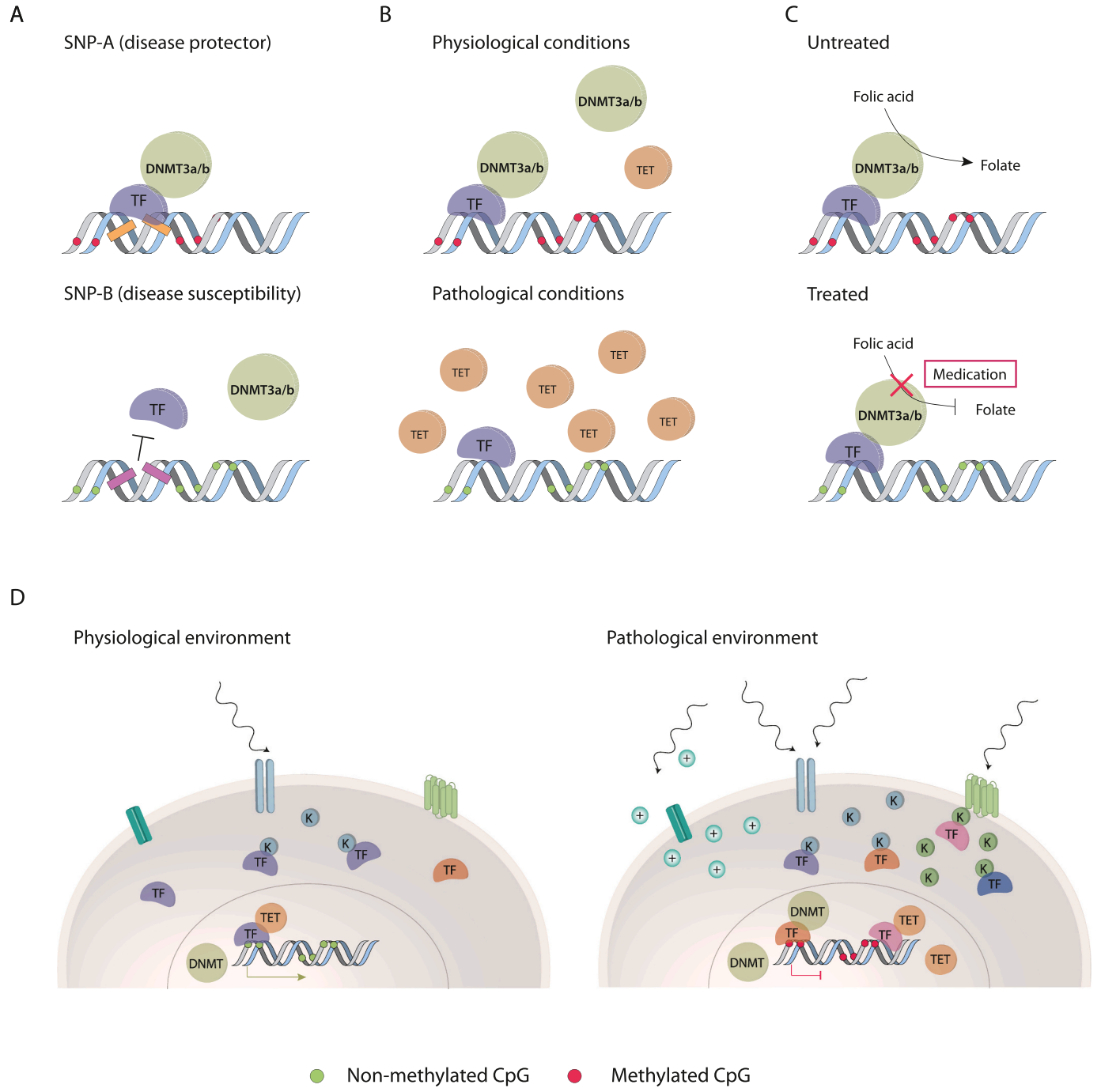


Fig. 2. Representative examples of DNA methylation alteration mechanisms in disease. A. SNP variation may have a role in DNA methylation through the differences in binding affinity of DNA binding proteins and DNA methylation machinery. B. Increase or decrease of DNA methylation-modifying enzyme proportions can lead to changes in the epigenomic profile. C. Some medications such as methotrexate may have a direct role in DNA methylation by inhibiting the enzymes responsible for regulating this modification. D. A variety of altered extracellular signals can lead to the activation of signaling pathways that drive different transcription factors or enzyme proportions and affinities.

Table 1
Summary of DNA methylation alterations in lupus.

Relation with disease	Cell type	Description	Highlighted genes	Ref
Susceptibility	Naïve CD4+ T cells	86 DMPs that primed the cells for a strong response to interferon.	*BST2, IFI44L, PARP9, STAT1*	[53]
	CD4+ T cells	*CD40LG* hypomethylation, leading to its overexpression and overproduction of IgG by B cells.	*CD40LG*	[63]
	Whole blood	Genetic regulation of 466 DMPs in SLE.	*IRF5, IRF7, UBE2L3*	[72]
	Whole blood	Methylation changes in *BANK1* are related to SNPs variations which are related to lupus nephritis.	*BANK1*	[73]
Classification and prognosis	Naïve CD4+ T cells	Different profiles between patients with discoid rash, malar rash and no cutaneous affectation.	*LY6G5C, RNF39, TNXB*	[74]
	Naïve CD4+ T cells	Different profile between patients	*IRF7, IFI44, PARP12*	[75]
	CD4+ T cells	with and without renal affectation,	*IRF1, IRF8, NRLP2*	[76]
	PBMCs	particularly in interferon-related genes.	*IFIT1, IFI44*	[77]
	PBMCs	Methylation profile correlated with disease symptomatology and severity.	*IFI44L, MX1, PARP9, EPSTI1, PDE7A*	[78]
Disease activity	CD4+ T cells, CD19+ B cells, monocytes, granulocytes	Bigger hypomethylation profile in patients with recent history of lupus flare, particularly in interferon-regulated genes.	*IFITM1, IFI44L, PARP9, MX1, IFITM1, STAT1, DTX3L*	[85]
	PBMCs	Ten times more differentially methylated positions in patients with SLEDAI >6.	*DTX3L, PARP9, MX1, IFI44L, IRF7*	[86]
	Naïve CD4+ T cells	Methylation levels correlated with disease activity.	*EZH2, IL4, IL5, IL13, PU.1*	[87]
	Neutrophils	Methylation profile remained stable across varying disease severities.	*GALNT18*	[94]
	CD4+ T cells, CD19+ B cells, CD14+ monocytes	No differences in interferon-related genes methylation across different disease states.		[51]
Response to treatment	Leukocytes	In rheumatoid arthritis, baseline methylation correlated with response to methotrexate treatment.		[98]
	Whole blood	In rheumatoid arthritis, baseline methylation did not correlate with response to methotrexate treatment.		[99]

3. DNA methylation in lupus classification and prognosis

SLE is an extremely heterogeneous disease with a wide variety of presentations in different individuals. It is composed by different pathophysiology, clinical manifestations, and progression trajectories, and these are generally difficult to anticipate. It is therefore inherently interesting to identify markers that can predict the evolution of the disease in individual patients. In the last decade, it has been repeatedly demonstrated that DNA methylation can potentially be used as a biomarker to differentiate between lupus subtypes and to predict disease severity.

In 2015, Renauer and colleagues [74] demonstrated that DNA methylation profiles from naïve CD4+ T cells differed significantly between lupus patients with and without a history of cutaneous manifestations. They also found differences between patients with a history of discoid rash and those with a history of malar rash. The most extensive differentially methylated region in discoid rash was hypermethylated and contained the leucocyte antigen-6 gene of the major histocompatibility (MHC) class III genomic region, *LY6G5C*. For malar rash, the most extensive region was also hypermethylated and encompassed *RNF39*, a gene of the MHC class I genomic region. In the non-cutaneous patients, the most extensive region was a hypomethylated section around the *TNXB* gene that encodes tenascin XB extracellular matrix protein, which is part of the MHC class III genomic region.

Several studies have focused on lupus nephritis, one of the most serious clinical manifestations in lupus patients. Renal biopsy is currently the standard technique for diagnosing this condition in affected patients. It is of great interest to identify additional markers for lupus nephritis that could potentially help clinicians diagnose and forecast the course of the disease. These studies have found differential DNA methylation patterns between patients with and without renal affectation, in peripheral blood mononuclear cells (PBMCs) and in naïve CD4+ T cells. The studies coincided in finding demethylation in genes involved in the interferon signaling pathway, e.g., *IRF7* [75], *IRF1* and *IRF8* [76], and *IFIT1* and *IFI44* [77].

Finally, in 2019, Lanata and colleagues published a very thorough study of genomic DNA methylation of PBMCs in 333 patients with lupus [78]. They classified the patients into three groups according to the severity of their disease: mild, severe 1 and severe 2 (the most severe subtype). The authors profiled the methylation status of >850,000 CpG positions across the genome and were able to identify 256 CpGs from 124 genes that had a differential methylation level depending on the clinical cluster. The annotated genes were particularly enriched in Type I interferon signaling, and in antiviral responses and inflammatory pathways. CpGs with the greatest variance across clusters were ascribed to the genes *IFI44L*, *MX1*, *PARP9*, *EPSTI1*, and *PDE7A*. They all displayed hypermethylation in the mild cluster relative to the severe clusters. Notably, most of them were interferon-responsive genes, indicating that the interferon signaling pathway is particularly frequently hypomethylated, thereby potentially yielding higher transcriptional levels, in the most severe forms of SLE. Unfortunately, it is undoubtedly difficult to determine whether this hypomethylation is a priming factor or a consequence of the elevated interferon levels found in the sera of lupus patients [79,80]. Nevertheless, hypomethylation of interferon-signaling genes is beyond doubt a marker for high disease activity in patients with SLE.

4. DNA methylation and lupus activity

Lupus, and other autoimmune diseases, are characterized by periods of high activity, known as flares, followed by periods of remission upon treatment, when the inflammatory status returns to basal/normal conditions and the clinical manifestations disappear. In order to stratify disease activity methodically, experts developed the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) in 1985 [81]. The index was updated in 2002 [82], and has been widely used in clinical [83,84]

and research [51,85] contexts.

DNA methylation changes in some genes and biological pathways are associated with disease activity in SLE. Again, the interferon-signaling pathway has been shown to have an important role in the pathogenesis of SLE because it is more aberrantly methylated in patients with higher activity indices. Firstly, in 2018, Jacobsen's group published some particularly significant findings that addressed this possibility [85]. They used 15 pairs of twins discordant for being affected by SLE to study the role DNA methylation in the disease. Discordant twins are an exceptionally useful cohort with which to study the effect of epigenetics in disease development since most of the variability introduced by genetic and environmental factors are reduced to a minimum, especially in identical twins. The researchers used these valuable samples to identify significant hypomethylation in interferon-regulated genes in all the cell types analyzed, i.e., CD4+ T cells, CD19+ B cells, monocytes, and granulocytes. The list of genes included *IFI44L*, *IFITM1*, and *PARP9*. Notably, hypomethylation was more pronounced in patients with a recent history of lupus flare. Secondly, a recent study in PBMCs examined 57 lupus patients with varying degrees of disease activity [86]. Separate comparisons between healthy donors and patients with SLEDAI $\leq$6 and between healthy donors and patients with SLEDAI >6 were made, revealing 6 and 79 significant differentially methylated positions (DMPs), respectively. In the SLEDAI $\leq$6 comparison, all the CpGs were hypomethylated in lupus patients and only one of them was ascribed to an interferon-related gene (*IRF2*). In the SLEDAI >6 comparison, 76% of the CpGs were hypomethylated and several of them were ascribed to interferon-related genes, including the two most differentially methylated CpGs, which were ascribed to the *PARP9* and *DTX3L* genes.

Other studies have found a link between disease activity and DNA methylation in other, interferon-independent pathways. This is the case for PU.1, a key transcription factor in myeloid and lymphoid lineage development, and some cytokine-encoding genes such as *IL4*, *IL5*, and *IL13*. Methylation at CpGs ascribed to these genes was found to be negatively correlated with disease activity in the first genome-wide study to investigate the association of DNA methylation with disease activity [87]. Coit and colleagues reported a positive correlation between DNA methylation and disease activity for the *EZH2* and *SUZ12* genes, which are components of the Polycomb repressive complex 2. Subsequently, another study demonstrated the important role of EZH2 in lupus development by describing its overexpression in lupus-naïve CD4+ T cells [88]. Furthermore, they overexpressed *EZH2* in healthy naïve CD4+ T cells to simulate *EZH2* strongly expressing lupus T cells and found genome-wide differences in DNA methylation upon *EZH2* expression. Genes such as *IL34* and *CTLA4*, as well as *EZH2* itself, exhibited differences in methylation under these conditions, highlighting the important role this factor may have in certain immune functions.

Genes encoding for other cytokines, like *IL-2* and *IL-10* have also been reported to be aberrantly regulated through DNA methylation in lupus samples, correlating with disease activity. On the one hand, IL-2 production is in part negatively regulated by protein phosphatase 2A (PP2Acα) [89] which is significantly hypomethylated in high activity lupus samples compared to low activity samples [90]. This leads to a decrease in IL-2 production in lupus patients that has effects on both promotion and suppression of a healthy immune response [91]. On the other hand, IL-10 expression has been described to be altered in lupus patients, in correlation with disease activity [92]. Its promoter and interferon-regulated response element regions showed reduced methylation levels in SLE T cells which lead to a reduction in IL-10 expression through impaired STAT3 and STAT5 recruitment to regulatory regions [93].

On the other hand, several studies have found no relation between DNA methylation levels and disease activity. This is the case of the study performed on the neutrophils of 54 lupus patients in a 4-year longitudinal study [94], which found that the DNA methylation profile of neutrophils generally remained stable over time and among disease states. Only two CpGs appeared to be correlated with disease activity, being ascribed to the *SNX18* and *FGD1* genes.

Counter to the results presented above, some studies found no significant correlation between interferon-regulated genes and disease activity. In 2013, Absher and colleagues published the results of a very extensive analysis of the DNA methylation profile of CD4+ T cells, CD19+ B cells, and CD14+ monocytes [51]. They found a commonly hypomethylated pattern of interferon-regulated genes across cell types. However, this hypomethylation was not associated with flare or quiescent states in their samples. Similar results were reported for *BST2* and *IFI44L*, two interferon-responsive genes [53]. Even though these were both hypomethylated and overexpressed in SLE CD4+ T cells, their methylation levels were stable across different disease states.

In summary, DNA methylation and SLE disease activity as measured by SLEDAI yield some apparently contradictory results. This could be due to different ranges of disease activity and the variety of analytical approaches used in the studies. For example, the higher the SLEDAI indices included in a study, the more the inflammatory and pathological environment that the sample probably contains. Inevitably, there is a higher probability of finding differences in high-activity samples than in mild-activity samples when compared with low-activity samples. The studies presented were based on different concepts of disease activity. Some considered patients' recent medical records [85] while others used only the disease state at the time of sample extraction. These approaches address different but equally interesting questions, but the inclusion of recent medical records as a covariate can be highly advantageous because it might make the analysis more sensitive to underlying inflammatory environments that are not reflected in a single SLEDAI measurement.

5. DNA methylation and response to treatment

Currently, SLE treatment options offer the possibility of proposing Lupus Low-Disease Activity State (LLDAS) as a conceivable goal for most patients [95,96]. The present recommendations for disease management establish a basal treatment with hydroxychloroquine and glucocorticoids (GCs). It is aimed to keep GC doses as low as possible or to withdraw them almost entirely to minimize the secondary effects of long-term administration. When lupus flares are not properly controlled by this basal treatment, the guidelines suggest the introduction of either immunosuppressive therapy like methotrexate, azathioprine, mycophenolate mofetil, or cyclophosphamide, or of biological therapies such as belimumab or rituximab [97]. Most of these therapies produce serious secondary effects and are not equally effective for all patients. Reliable biomarkers that predict the response to treatment are required in order to direct patients to their most suitable therapy. DNA methylation could potentially assume this role since it can offer a stable profile of the inflammatory pathways affected in each patient. However, no results from any studies have so far been published regarding this, although it is notable that there are some publications about DNA methylation predicting the response to treatment with two of these drugs, methotrexate and rituximab, in the context of other diseases.

In rheumatoid arthritis, another autoimmune disease, two recent studies established that DNA methylation profiles differed between patients who responded or did not respond to methotrexate. In whole-blood leukocytes, higher global baseline DNA methylation levels were associated with higher levels of disease activity and with methotrexate non-response 3 months after treatment [98]. These results were contradicted by a subsequent study in which no difference in baseline methylation was found between the responder and non-responder groups. However, at the individual CpG level, 4 weeks after treatment, the two groups displayed differentially methylated positions. These CpG positions mapped to *GATA3*, an important factor in Th2 differentiation, and *mir-182*, a microRNA relevant to osteoclast differentiation [99].

Many of the therapies used for SLE are also used to treat certain cancers. Although the mechanism of action probably differs between

these groups of pathologies, two studies of DNA methylation and response to treatment have been published in cancer that may be relevant to SLE. For example, the aberrant methylation of the promoter of the *RFC* gene, which is responsible for the transport of folic acid, was found to be associated with resistance to methotrexate treatment in central nervous system lymphomas [100]. Similarly, methylation of *p15* in diffuse large B-cell lymphoma, in combination with gender and tumor size, was identified as a predictive factor for response to rituximab [101].

It is worth pointing out that several of the drugs intended for use in SLE treatment have effects on DNA methylation when administered. For instance, GC-treated lupus patients presented lower global DNA methylation levels than non-treated patients [102]. In the case of mycophenolate, *IFNG* methylation levels increased in naïve CD4+ T cells after treatment and stimulation. In other words, stimulation of naïve CD4+ T cells induced a reduction in DNA methylation that was reversed with mycophenolate treatment. The treatment did not have this effect on memory CD4+ T cells [103]. Neither global DNA methylation levels nor the promoter region of *CD40L* were affected in lupus CD4+ T cells upon treatment with mycophenolate. However, this region was epigenetically modified by mycophenolate with respect to histone H4 acetylation and histone H3 lysine 3 methylation levels [104].

The mechanisms of action of some of these drugs may be invalidated by aberrant DNA methylation patterns. This has been described in some small cell lung cancer cells in which GC receptor genes are hypermethylated, resulting in a loss of expression, which causes these cells to be resistant to GC treatment. This defect can be easily corrected in vitro with demethylating agents, rendering the cells sensitive to GCs, and confirming the DNA methylation-driven resistance to treatment [105]. Equivalent results were observed in three other cell lines described as being resistant to GC treatment through this mechanism [106].

To date, little evidence has emerged of the existence of response-specific DNA methylation profiles in SLE or that supports their use as predictive biomarkers. However, based on the results that link the most widely used treatments for SLE with several DNA methylation aberrations, it seems to be a very promising field that is likely to yield interesting results if pursued further. Prospective findings could easily become relevant in treatment decisions, given the great heterogeneity in SLE pathophysiology and response to treatments.

Finally, epigenetic alterations present in lupus, such as those presented in this review, could be translated into potential targets for the development of treatments, in a similar way to some cancer therapies [107]. So far, all the DNA methylation-targeting FDA-approved drugs target DNA hypermethylation. Given the extended hypomethylation present in lupus patients, particularly in the interferon-signaling pathway, it could be interesting to develop drugs targeting hypomethylation. S-Adenosylmethione (SAM) is the donor of the methyl group in the reactions catalyzed by DNMTs to incorporate DNA methylation. Several attempts to modify DNA hypomethylation have been made by modifying SAM levels. These are in part dependent on diet components such as folate and methionine [108]. Richardson's group already described in 2010 that restriction of these micronutrients in vitro could revert typical lupus-like gene expression in T cells from older donors [109]. Other groups have attempted to treat DNA hypomethylation-mediated abnormalities with the supplementation of SAM in cell lines and animal models (reviewed in [110]). This has supposed a reduction of tumor size by increasing apoptotic cell death and reducing cell proliferation. Interestingly, SAM treatment also improved the outcome of inflammation-induced color cancer by modulating pro-inflammatory pathways [111]. These results point out at DNA methylation as a new target for modulating the immune system in diseases with aberrant DNA methylation profiles such as lupus.

6. How can we improve DNA methylation studies in lupus?

As set out in this review, the field of lupus research has undergone clinically significant developments in relation to the identification of DNA methylation alterations. However, many questions remain unanswered and the need for accurate biomarkers that can be used in the clinics is unmet. Reliable biomarkers are constantly required at all stages in the course of this disease: susceptibility, diagnosis, development, and response to treatment. We propose that DNA methylation is an excellent candidate biomarker for two main reasons: first, DNA methylation is associated with many clinical aspects in lupus and, second, methylation of cytosines is a relatively stable chemical modification, and so it can be more reliably measured than other factors, such as transcriptional profiles and those related to immune cells, for example, cytometry and cytokine levels. Nevertheless, several improvements in the design of these studies could be made that would produce more trustworthy and applicable results.

DNA methylation profiles are cell type-specific. Reports show that up to 63% of the CpG sites examined in an epigenome-wide analysis of blood samples differed significantly among cell types [112]. Therefore, not considering cell composition when comparing individuals could lead to erroneous conclusions. Although it is possible to partially correct for this confounding issue in the subsequent bioinformatic analysis [112–115], a better approach would be to profile each immune cell type separately. Furthermore, differences in DNA methylation are present at the single-cell level [116] and several techniques for single-cell methylome analysis have been developed [117]. However, no such techniques have so far been applied in studies of DNA methylation in lupus.

Another factor to consider when studying epigenetic profiles is the cohort size and any confounding variables. DNA methylation is influenced by lifestyle and environmental factors such as diet, physical activity, environmental pollutants, and tobacco smoking [118]. These are often not considered when recruiting individuals for research studies and can introduce bias into the results. Therefore, it is particularly important when studying elements that are as strongly influenced by environmental conditions as is DNA methylation to increase the size of the cohort and account for as many potential confounding variables as possible, in order to minimize this involuntary sampling bias.

Contradictory or inconclusive results may also arise from not considering disease activity and other clinical parameters, such as damage, that have been used in some studies. Several of those mentioned in this review have conclusively shown that differences in DNA methylation are more pronounced in patients with active disease than in those in remission. Consequently, studying patients with higher activity scores makes it more likely that differences due to disease activity will be identified. On the other hand, several studies have reported that current clinical and serological evaluation of SLE may not be an entirely accurate approach to assessing the inflammatory state of patients. For example, after 12 months of clinical remission following a renal flare, renal biopsy showed that 44% of patients still had histological activity in the kidney [119]. This sustained inflammatory state could not be detected from the clinical parameters and the activity index of the patients was low. For these reasons, it is worthwhile including different elements of the recent clinical history of patients as covariates in studies of lupus methylation.

7. Conclusions

Most of the information about DNA methylation profiles in lupus points towards a propensity associated with interferon-regulated genes. Hypothetically, DNA methylation alterations in these genes could increase the predisposition of lupus immune cells to respond to interferon signaling. The role of this cytokine in the pathogenesis of lupus has been widely demonstrated, as well as its correlation with the genetic defects in the highly genetic forms of childhood-onset lupus. As recently reviewed in [120], genetic variance and overexpression of Type I interferon are both associated with SLE. Moreover, its dysregulation has been linked not only to the initial break in tolerance but also to the propagation of the ongoing disease. For these reasons, Type I interferon

has been recognized as a key pathogenic factor in SLE and proposed as a therapeutic target [121].

Similar to the identified epigenetic dysregulation associated with interferon unbalance, other key players in this disease may be discovered from the study of DNA methylation. So far, most DNA methylation studies have been performed on adult patients. The inclusion of childhood-onset patients in the studies could help elucidate the mechanisms underlying this typically more severe form of the disease. SLE is a complex autoimmune disease with a wide variety of manifestations and severities. Defining disease biomarkers that can help clinicians understand, predict, diagnose, and treat the different stages of this disease is crucial if we are to realize our aim of permanent remission. DNA methylation of immune system cells has been proposed as being able to act in this way because of consistent lines of evidence, documented in this review, linking DNA methylation and SLE pathogenesis. Other biomarkers in clinical practice include genetic variance, gene expression, and protein levels. However, these are of limited value because they are overly stable or overly variable, respectively. DNA methylation could overcome these constraints because of its semi-stable profile. Moreover, as the profiles presented are of immune system cells, the low-invasive nature of the methods used to obtain them makes this approach a very attractive means of generating biomarkers.

Nevertheless, more results, of a non-conflicting nature, are required. In this review, we propose several ways to obtain clearer results than the mixed findings obtained so far. Considering covariates and patients' clinical history is very easy to implement in routine practice and could make a difference to the significance and reliability of the results obtained. Furthermore, the separation of cell types before studying DNA methylation could reveal differences related to the variation in population proportions. Although this process is not as easily implemented in clinical practice with current technologies, the results obtained by this enhancement could help us better understand the pathogenesis of the disease. It could become part of standard clinical analysis with future technological advancements of cell separation or DNA methylation studies such as single-cell DNA methylation profiling.

Acknowledgements

We thank the CERCA Programme/Generalitat de Catalunya and the Josep Carreras Foundation for institutional support. E.B. was funded by the Spanish Ministry of Science and Innovation (MICINN; grant number PID2020-117212RB-I00; AEI / 10.13038/501100011033/).

References

[1] M. Aringer, K. Costenbader, D. Daikh, R. Brinks, M. Mosca, et al., European league against rheumatism/American College of rheumatology classification criteria for systemic lupus erythematosus, Arthritis Rheumatol. 71 (2019) (2019) 1400–1412.
[2] M.D. Alonso, J. Llorca, F. Martinez-Vazquez, J.A. Miranda-Filloy, T. Diaz De Teran, et al., Systemic lupus erythematosus in northwestern Spain: a 20-year epidemiologic study, Medicine (Baltimore) 90 (2011) 350–358.
[3] G.J. Pons-Estel, G.S. Alarcón, L. Scofield, L. Reinlib, G.S. Cooper, Understanding the epidemiology and progression of systemic lupus erythematosus, Semin. Arthritis Rheum. 39 (2010) 257–268.
[4] C.H. Feldman, L.T. Hiraki, J. Liu, M.A. Fischer, D.H. Solomon, et al., Epidemiology and sociodemographics of systemic lupus erythematosus and lupus nephritis among US adults with Medicaid coverage, 2000-2004, Arthritis Rheum. 65 (2013) 753–763.
[5] S.S. Lim, A.R. Bayakly, C.G. Helmick, C. Gordon, K.A. Easley, et al., The incidence and prevalence of systemic lupus erythematosus, 2002-2004: the Georgia lupus registry, Arthritis Rheumatol. 66 (2014) 357–368.
[6] M.S. Lo, Insights gained from the study of pediatric systemic lupus erythematosus, Front. Immunol. 9 (2018).
[7] Y.F. Wang, Y. Zhang, Z. Lin, H. Zhang, T.Y. Wang, et al., Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups, Nat. Commun. 12 (2021).
[8] J. Bentham, D.L. Morris, D.S. Cunninghame Graham, C.L. Pinder, P. Tombleson, et al., Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus, Nat. Genet. 47 (2015) 1457–1464.
[9] S.V. Kozyrev, A.K. Abelson, J. Wojcik, A. Zaghlool, M.V.P. Linga Reddy, et al., Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus, Nat. Genet. 40 (2008) 211–216.
[10] J.E. Molineros, A.K. Maiti, C. Sun, L.L. Looger, S. Han, et al., Admixture mapping in lupus identifies multiple functional variants within IFIH1 associated with apoptosis, inflammation, and autoantibody production, PLoS Genet. 9 (2013).
[11] V. Gateva, J.K. Sandling, G. Hom, K.E. Taylor, S.A. Chung, et al., A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus, Nat. Genet. 41 (2009) 1228–1233.
[12] C. Kyogoku, C.D. Langefeld, W.A. Ortmann, A. Lee, S. Selby, et al., Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE, Am. J. Hum. Genet. 75 (2004) 504–507.
[13] B. Namjou, X. Kim-Howard, C. Sun, A. Adler, S.A. Chung, et al., PTPN22 Association in Systemic Lupus Erythematosus (SLE) with respect to individual ancestry and clinical sub-phenotypes, PLoS One 8 (2013).
[14] W. Yang, H. Tang, Y. Zhang, X. Tang, J. Zhang, et al., Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians, Am. J. Hum. Genet. 92 (2013) 41–51.
[15] E. Sánchez, M.E. Comeau, B.I. Freedman, J.A. Kelly, K.M. Kaufman, et al., Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study, Arthritis Rheum. 63 (2011) 3493–3501.
[16] J.W. Han, H.F. Zheng, Y. Cui, L.D. Sun, D.Q. Ye, et al., Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus, Nat. Genet. 41 (2009) 1234–1237.
[17] S.K. Nath, S. Han, X. Kim-Howard, J.A. Kelly, P. Viswanathan, et al., A nonsynonymous functional variant in integrin-αM (encoded by ITGAM) is associated with systemic lupus erythematosus, Nat. Genet. 40 (2008) 152–154.
[18] K. Kim, E.E. Brown, C.B. Choi, M.E. Alarcón-Riquelme, J.A. Kelly, et al., Variation in the ICAM1-ICAM4-ICAM5 locus is associated with systemic lupus erythematosus susceptibility in multiple ancestries, Ann. Rheum. Dis. 71 (2012) 1809–1814.
[19] R. Webb, J.A. Kelly, E.C. Somers, T. Hughes, K.M. Kaufman, et al., Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients, Ann. Rheum. Dis. 70 (2011) 151–156.
[20] Y. Bin Joo, J. Lim, B.P. Tsao, S.K. Nath, K. Kim, et al., Genetic variants in systemic lupus erythematosus susceptibility loci, XKR6 and GLT1D1 are associated with childhood-onset SLE in a Korean cohort, Sci. Rep. 8 (2018).
[21] D. Webber, J. Cao, D. Dominguez, D.D. Gladman, D.M. Levy, et al., Association of systemic lupus erythematosus (SLE) genetic susceptibility loci with lupus nephritis in childhood-onset and adult-onset SLE, Rheumatology (Oxford) 59 (2020) 90–98.
[22] A. Belot, G.I. Rice, S.O. Omarjee, Q. Rouchon, E.M.D. Smith, et al., Contribution of rare and predicted pathogenic gene variants to childhood-onset lupus: a large, genetic panel analysis of British and French cohorts, Lancet Rheumatol. 2 (2020) e99–e109.
[23] M.J. Walport, K.A. Davies, B.J. Morley, M. Botto, Complement deficiency and autoimmunity, Ann. N. Y. Acad. Sci. 815 (1997) 267–281.
[24] A.R. Bryan, E.Y. Wu, Complement deficiencies in systemic lupus erythematosus, Curr Allergy Asthma Rep 14 (2014).
[25] S.M. Al-Mayouf, A. Sunker, R. Abdwani, S. Al Abrawi, F. Almurshedi, et al., Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus, Nat. Genet. 43 (2011) 1186–1188.
[26] M.A. Lee-Kirsch, M. Gong, D. Chowdhury, L. Senenko, K. Engel, et al., Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 are associated with systemic lupus erythematosus, Nat. Genet. 39 (2007) 1065–1067.
[27] M. Barbhaiya, K.H. Costenbader, Environmental exposures and the development of systemic lupus erythematosus, Curr. Opin. Rheumatol. 28 (2016) 497–505.
[28] D. Deafen, A. Escalante, L. Weinrib, D. Horwitz, B. Bachman, et al., A revised estimate of twin concordance in systemic lupus erythematosus, Arthritis Rheum. 35 (1992) 311–318.
[29] C.M. Brunner, D.A. Horwitz, M.K. Shann, B.A. Sturgill, J.S. Davis IV, Clinical and immunologic studies in identical twins discordant for systemic lupus erythematosus, Am. J. Med. 55 (1973) 249–254.
[30] S. Bustabad, T. González, E. Trujillo, Systemic lupus erythematosus in childhood. Concordance in a pair of monozygotic twins for anti-dsDNA antibodies and discordance for disease expression, J. Rheumatol. 24 (1997) 1450–1451.
[31] K.H. Costenbader, D.J. Kim, J. Peerzada, S. Lockman, D. Nobles-Knight, et al., Cigarette smoking and the risk of systemic lupus erythematosus: a meta-analysis, Arthritis Rheum. 50 (2004) 849–857.
[32] M. Barbhaiya, S.K. Tedeschi, B. Lu, S. Malspeis, D. Kreps, et al., Cigarette smoking and the risk of systemic lupus erythematosus, overall and by anti-double stranded DNA antibody subtype, in the nurses' health study cohorts, Ann. Rheum. Dis. 77 (2018) 196.
[33] J. Sanchez-Guerrero, E.W. Karlson, M.H. Liang, D.J. Hunter, F.E. Speizer, et al., Past use of oral contraceptives and the risk of developing systemic lupus erythematosus, Arthritis Rheum. 40 (1997) 804–808.
[34] K.H. Costenbader, D. Feskanich, M.J. Stampfer, E.W. Karlson, Reproductive and menopausal factors and risk of systemic lupus erythematosus in women, Arthritis Rheum. 56 (2007) 1251–1262.
[35] P.R. Nielsen, T.W. Kragstrup, B.W. Deleuran, M.E. Benros, Infections as risk factor for autoimmune diseases - a nationwide study, J. Autoimmun. 74 (2016) 176–181.

[36] C.G. Parks, A. de Souza Espindola, M. Santos, K.H. Costenbader Barbhaiya, Understanding the role of environmental factors in the development of systemic lupus erythematosus, Best Pract. Res. Clin. Rheumatol. 31 (2017) 306–320.

[37] A. Charras, E. Smith, C.M. Hedrich, Systemic lupus erythematosus in children and young people, Curr. Rheumatol. Rep. 23 (2021).

[38] G. Gulati, H.I. Brunner, Environmental triggers in systemic lupus erythematosus, Semin, Arthritis Rheum. 47 (2018) 710–717.

[39] M. Ehrlich, DNA hypermethylation in disease: mechanisms and clinical relevance, Epigenetics. 14 (2019) 1141–1163.

[40] Y. Pan, G. Liu, F. Zhou, B. Su, Y. Li, DNA methylation profiles in cancer diagnosis and therapeutics, Clin. Exp. Med. 18 (2018) 1–14.

[41] B. Richardson, L. Scheinbart, J. Strahler, L. Gross, S. Hanash, et al., Evidence for impaired t cell dna methylation in systemic lupus erythematosus and rheumatoid arthritis, Arthritis Rheum. 33 (1990) 1665–1673.

[42] B.M. Javierre, A.F. Fernandez, J. Richter, F. Al-Shahrour, J. Ignacio Martin-Subero, et al., Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus, Genome Res. 20 (2010) 170–179.

[43] C.R. Keenan, R.S. Allan, Epigenomic drivers of immune dysfunction in aging, Aging Cell 18 (2019).

[44] M.G. Dozmorov, P. Coit, K. Maksimowicz-Mckinnon, A.H. Sawalha, Age-Associated DNA Methylation Changes in naive CD4+ T Cells Suggest an Evolving Autoimmune Epigenotype in Aging T Cells 9, 2017, pp. 429–445, https://doi.org/10.2217/Epi-2016-0143.

[45] Z. Zhang, C. Deng, Q. Lu, B. Richardson, Age-dependent DNA methylation changes in the ITGAL (CD11a) promoter, Mech. Ageing Dev. 123 (2002) 1257–1268.

[46] M.A. Jeffries, M. Dozmorov, Y. Tang, J.T. Merrill, J.D. Wren, et al., Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus, Epigenetics 6 (2011) 593–601.

[47] P.G. Bronson, C. Chaivorapol, W. Ortmann, T.W. Behrens, R.R. Graham, The genetics of type I interferon in systemic lupus erythematosus, Curr. Opin. Immunol. 24 (2012) 530–537.

[48] E.C. Baechler, F.M. Batliwalla, G. Karypis, P.M. Gaffney, W.A. Ortmann, et al., Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 2610–2615.

[49] D.M. Santer, B.E. Hall, T.C. George, S. Tangsombatvisit, C.L. Liu, et al., C1q deficiency leads to the defective suppression of IFN-alpha in response to nucleoprotein containing immune complexes, J. Immunol. 185 (2010) 4738–4749.

[50] P. Coit, S. Yalavarthi, M. Ognenovski, W. Zhao, S. Hasni, et al., Epigenome profiling reveals significant DNA demethylation of interferon signature genes in lupus neutrophils, J. Autoimmun. 58 (2015) 59–66.

[51] D.M. Absher, X. Li, L.L. Waite, A. Gibson, K. Roberts, et al., Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations, PLoS Genet. 9 (2013).

[52] S. Miller, P.S. Tsou, P. Coit, E. Gensterblum-Miller, P. Renauer, et al., Hypomethylation of STAT1 and HLA-DRB1 is associated with type-I interferon-dependent HLA-DRB1 expression in lupus CD8+ T cells, Ann. Rheum. Dis. 78 (2019) 519–528.

[53] P. Coit, M. Jeffries, N. Altorok, M.G. Dozmorov, K.A. Koelsch, et al., Genome-wide DNA methylation study suggests epigenetic accessibility andtranscriptional poising of interferon-regulated genes in naïve CD4+ T cellsfrom lupus patients, J. Autoimmun. 43 (2013) 78–84.

[54] A.H. Sawalha, L. Wang, A. Nadig, E.C. Somers, W.J. McCune, et al., Sex-specific differences in the relationship between genetic susceptibility, T cell DNA demethylation and lupus flare severity, J. Autoimmun. 38 (2012) J216–J222.

[55] L.J. Lu, D.J. Wallace, M.L. Ishimori, R.H. Scofield, M.H. Weisman, Review: male systemic lupus erythematosus: a review of sex disparities in this disease, Lupus 19 (2010) 119–129.

[56] X. Huang, N. Jia, F. Xiao, C. Sun, J. Zhu, et al., Differences in the clinical manifestations and mortality of systemic lupus erythematosus onset in children and adults: a systematic review and meta-analysis, Int. Arch. Allergy Immunol. 183 (2021) 116–126. In press.

[57] C. Selmi, The X in sex: how autoimmune diseases revolve around sex chromosomes, Best Pract. Res. Clin. Rheumatol. 22 (2008) 913–922.

[58] R.H. Scofield, G.R. Bruner, B. Namjou, R.P. Kimberly, R. Ramsey-Goldman, et al., Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: support for the notion of a gene-dose effect from the X chromosome, Arthritis Rheum. 58 (2008) 2511–2517.

[59] K. Liu, B.T. Kurien, S.L. Zimmerman, K.M. Kaufman, D.H. Taft, et al., X chromosome dose and sex bias in autoimmune diseases: increased prevalence of 47,XXX in systemic lupus erythematosus and Sjögren's syndrome, Arthritis Rheumatol. 68 (2016) 1290–1300.

[60] C.M. Cooney, G.R. Bruner, T. Aberle, B. Namjou-Khales, L.K. Myers, et al., 46,X, del(X)(q13) Turner's syndrome women with systemic lupus erythematosus in a pedigree multiplex for SLE, Genes Immun. 10 (2009) 478–481.

[61] M.T. Ross, D.V. Grafham, A.J. Coffey, S. Scherer, K. McLay, et al., The DNA sequence of the human X chromosome, Nature. 434 (2005) 325–337.

[62] A. Hewagama, G. Gorelik, D. Patel, P. Liyanarachchi, McCune WJ, et al., Overexpression of X-linked genes in T cells from women with lupus, J. Autoimmun. 41 (2013) 60–71.

[63] Q. Lu, A. Wu, L. Tesmer, D. Ray, N. Yousif, et al., Demethylation of CD40LG on the inactive X in T cells from women with lupus, J. Immunol. 179 (2007) 6352–6358.

[64] Y. Zhou, J. Yuan, Y. Pan, Y. Fei, X. Qiu, et al., T cell CD40LG gene expression and the production of IgG by autologous B cells in systemic lupus erythematosus, Clin. Immunol. 132 (2009) 362–370.

[65] P. Enghard, J.Y. Humrich, B. Rudolph, S. Rosenberger, R. Biesen, et al., CXCR3+ CD4+ T cells are enriched in inflamed kidneys and urine and provide a new biomarker for acute nephritis flares in systemic lupus erythematosus patients, Arthritis Rheum. 60 (2009) 199–206.

[66] D. Wu, Y. Cai, J. Jin, Potential coordination role between O-GlcNAcylation and epigenetics, Protein Cell 8 (2017) 713–723.

[67] A. Golks, T.T.T. Tran, J.F. Goetschy, D. Guerini, Requirement for O-linked N-acetylglucosaminyltransferase in lymphocytes activation, EMBO J. 26 (2007) 4368–4379.

[68] J. Van Dongen, M.G. Nivard, G. Willemsen, J.J. Hottenga, Q. Helmer, et al., Genetic and environmental influences interact with age and sex in shaping the human methylome, Nat. Commun. 7 (2016).

[69] J.L. McClay, A.A. Shabalin, M.G. Dozmorov, D.E. Adkins, G. Kumar, et al., High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction, Genome Biol. 16 (2015).

[70] L.T. Husquin, M. Rotival, M. Fagny, H. Quach, N. Zidane, et al., Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation 06 biological sciences 0604 genetics, Genome Biol. 19 (2018).

[71] G.N. Goulielmos, M.I. Zervou, V.M. Vazgiourakis, Y. Ghodke-Puranik, A. Garyfallos, et al., The genetics and molecular pathogenesis of systemic lupus erythematosus (SLE) in populations of different ancestry, Gene 668 (2018) 59–72.

[72] J. Imgenberg-Kreuz, J.C. Almlöf, D. Leonard, A. Alexsson, G. Nordmark, et al., DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus, Ann. Rheum. Dis. 77 (2018) 736–743.

[73] K. Bolin, J. Imgenberg-Kreuz, D. Leonard, J.K. Sandling, A. Alexsson, et al., Variants in BANK1 are associated with lupus nephritis of European ancestry, Genes Immun. 22 (2021) 194–202.

[74] P. Renauer, P. Coit, M.A. Jeffries, J.T. Merrill, W.J. McCune, et al., DNA methylation patterns in naïve CD4+ T cells identify epigenetic susceptibility loci for malar rash and discoid rash in systemic lupus erythematosus, Lupus Sci. Med. 2 (2015).

[75] P. Coit, P. Renauer, M.A. Jeffries, J.T. Merrill, W.J. McCune, et al., Renal involvement in lupus is characterized by unique DNA methylation changes in naïve CD4+ T cells, J. Autoimmun. 61 (2015) 29–35.

[76] M. Zhao, S. Liu, S. Luo, H. Wu, M. Tang, et al., DNA methylation and mRNA and microRNA expression of SLE CD4+ T cells correlate with disease phenotype, J. Autoimmun. 54 (2014) 127–136.

[77] A. Mok, O. Solomon, R.R. Nayak, P. Coit, H.L. Quach, et al., Genome-wide profiling identifies associations between lupus nephritis and differential methylation of genes regulating tissue hypoxia and type 1 interferon responses, Lupus Sci. Med. 3 (2016).

[78] C.M. Lanata, I. Paranjpe, J. Nititham, K.E. Taylor, M. Gianfrancesco, et al., A phenotypic and genomics approach in a multi-ethnic cohort to subtype systemic lupus erythematosus, Nat. Commun. 10 (2019).

[79] J.J. Hooks, H.M. Moutsopoulos, S.A. Geis, N.I. Stahl, J.L. Decker, et al., Immune interferon in the circulation of patients with autoimmune disease, N. Engl. J. Med. 301 (1979) 5–8.

[80] S.R. Ytterberg, T.J. Schnitzer, Serum interferon levels in patients with systemic lupus erythematosus, Arthritis Rheum. 25 (1982) 401–406.

[81] C. Bombardier, D.D. Gladman, M.B. Urowitz, D. Caron, C.H. Chang, et al., Derivation of the sledai. A disease activity index for lupus patients, Arthritis Rheum. 35 (1992) 630–640.

[82] D.D. Gladman, D. Ibañez, M.B. Urowltz, Systemic lupus erythematosus disease activity index 2000, J. Rheumatol. 29 (2002) 288–291.

[83] I. Parodis, S. Emamikia, A. Gomez, I. Gunnarsson, R.F. van Vollenhoven, et al., Clinical SLEDAI-2K zero may be a pragmatic outcome measure in SLE studies, Expert. Opin. Biol. Ther. 19 (2019) 157–168.

[84] J. Feld, D. Isenberg, Why and how should we measure disease activity and damage in lupus? Press. Medicale. 43 (2014).

[85] C.J. Ulff-Møller, F. Asmar, Y. Liu, A.J. Svendsen, F. Busato, et al., Twin DNA methylation profiling reveals flare-dependent interferon signature and B cell promoter hypermethylation in systemic lupus erythematosus, Arthritis Rheumatol. 70 (2018) 878–890.

[86] S. Joseph, N.I. George, B. Green-Knox, E.L. Treadwell, B. Word, et al., Epigenome-wide association study of peripheral blood mononuclear cells in systemic lupus erythematosus: identifying DNA methylation signatures associated with interferon-related genes based on ethnicity and SLEDAI, J. Autoimmun. 96 (2019) 147–157.

[87] P. Coit, M.G. Dozmorov, J.T. Merrill, W.J. McCune, K. Maksimowicz-McKinnon, et al., Epigenetic reprogramming in naive CD4+ T cells favoring T cell activation and non-Th1 effector T cell immune response as an early event in lupus flares, Arthritis Rheumatol. 68 (2016) 2200–2209.

[88] P.S. Tsou, P. Coit, N.C. Kilian, A.H. Sawalha, EZH2 modulates the DNA Methylome and controls T cell adhesion through junctional adhesion molecule a in lupus patients, Arthritis Rheumatol. 70 (2018) 98–108.

[89] C.G. Katsiari, V.C. Kyttaris, Y.T. Juang, G.C. Tsokos, Protein phosphatase 2A is a negative regulator of IL-2 production in patients with systemic lupus erythematosus, J. Clin. Invest. 115 (2005) 3193–3204.

[90] K. Sunahori, Y.-T. Juang, V.C. Kyttaris, G.C. Tsokos, Promoter hypomethylation results in increased expression of protein phosphatase 2A in T cells from patients with systemic lupus erythematosus, J. Immunol. 186 (2011) 4508–4517.

[91] A.G.A. Kolios, G.C. Tsokos, D. Klatzmann, Interleukin-2 and regulatory T cells in rheumatic diseases, Nat. Rev. Rheumatol. 17 (2021) 749–766.
[92] J. Godsell, I. Rudloff, R. Kandane-Rathnayake, A. Hoi, M.F. Nold, et al., Clinical associations of IL-10 and IL-37 in systemic lupus erythematosus, Sci. Rep. 6 (2016).
[93] C.M. Hedrich, T. Rauen, S.A. Apostolidis, A.P. Grammatikos, N.R. Rodriguez, et al., Stat3 promotes IL-10 expression in lupus T cells through trans-activation and chromatin remodeling, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 13457–13462.
[94] P. Coit, L. Ortiz-Fernandez, E.E. Lewis, W.J. McCune, K. Maksimowicz-McKinnon, et al., A longitudinal and transancestral analysis of DNA methylation patterns and disease activity in lupus patients, JCI Insight. 5 (2020).
[95] F. Basta, F. Fasola, K. Triantafyllias, A. Schwarting, Systemic lupus erythematosus (SLE) therapy: the old and the new, Rheumatol. Ther. 7 (2020) 433–446.
[96] C. Tani, R. Vagelli, C. Stagnaro, L. Carli, M. Mosca, Remission and low disease activity in systemic lupus erythematosus: an achievable goal even with fewer steroids? Real-life data from a monocentric cohort, Lupus Sci. Med. 5 (2018).
[97] A. Fanouriakis, M. Kostopoulou, A. Alunno, M. Aringer, I. Bajema, et al., Update of the EULAR recommendations for the management of systemic lupus erythematosus, Ann. Rheum. Dis. 78 (2019) (2019) 736–745.
[98] H.R. Gosselt, B.D. Van Zelst, M.C.F.J. De Rotte, J.M.W. Hazes, R. De Jonge, et al., Higher baseline global leukocyte DNA methylation is associated with MTX non-response in early RA patients, Arthritis Res. Ther. 21 (2019).
[99] N. Nair, D. Plant, D. Plant, S.M. Verstappen, S.M. Verstappen, et al., Differential DNA methylation correlates with response to methotrexate in rheumatoid arthritis, Rheumatol. (United Kingdom). 59 (2020) 1364–1371.
[100] A.J.M. Ferreri, S. Dell'Oro, D. Capello, M. Ponzoni, P. Iuzzolino, et al., Aberrant methylation in the promoter region of the reduced folate carrier gene is a potential mechanism of resistance to methotrexate in primary central nervous system lymphomas, Br. J. Haematol. 126 (2004) 657–664.
[101] M. Krajnović, M.P. Jovanović, B. Mihaljević, B. Andelić, O. Tarabar, et al., Hypermethylation of p15 gene in diffuse - large B-cell lymphoma: association with less aggressiveness of the disease, Clin. Transl. Sci. 7 (2014) 384–390.
[102] G.S. Wang, M. Zhang, X.P. Li, H. Zhang, W. Chen, et al., Ultraviolet B exposure of peripheral blood mononuclear cells of patients with systemic lupus erythematosus inhibits DNA methylation, Lupus 18 (2009) 1037–1044.
[103] F.S. Peters, A.M.A. Peeters, L.J. Hofland, M.G.H. Betjes, K. Boer, et al., Interferon-gamma DNA methylation is affected by mycophenolic acid but not by tacrolimus after T-cell activation, Front. Immunol. 8 (2017).
[104] Y. Yang, Q. Tang, M. Zhao, G. Liang, H. Wu, et al., The effect of mycophenolic acid on epigenetic modifications in lupus CD4+T cells, Clin. Immunol. 158 (2015) 67–76.
[105] P. Kay, G. Schlossmacher, L. Matthews, P. Sommer, D. Singh, et al., Loss of glucocorticoid receptor expression by DNA methylation prevents glucocorticoid induced apoptosis in human small cell lung cancer cells, PLoS One 6 (2011).
[106] A.L. Miller, C. Geng, G. Golovko, M. Sharma, J.R. Schwartz, et al., Epigenetic alteration by DNA-demethylating treatment restores apoptotic response to glucocorticoids in dexamethasone-resistant human malignant lymphoid cells, Cancer Cell Int. 14 (2014).
[107] Y. Cheng, C. He, M. Wang, X. Ma, F. Mo, et al., Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials, Signal Transduct. Target. Ther. 4 (2019).
[108] S.A. Ross, L. Poirier, Proceedings of the Trans-HHS Workshop: diet, DNA methylation processes and health, J. Nutr. 132 (2002).
[109] Y.P. Li, Y. Liu, F.M. Strickland, B. Richardson, Age-dependent decreases in DNA methyltransferase levels and low transmethylation micronutrient levels synergize to promote overexpression of genes implicated in autoimmunity and acute coronary syndromes, Exp. Gerontol. 45 (2010) 312–322.
[110] N. Mahmood, S.A. Rabbani, Targeting DNA hypomethylation in malignancy by epigenetic therapies, in: Adv. Exp. Med. Biol., Adv Exp Med Biol, 2019, pp. 179–196.
[111] T.W.H. Li, H. Yang, H. Peng, M. Xia, J.M. Mato, et al., Effects of S-adenosylmethionine and methylthioadenosine on inflammation-induced colon cancer in mice, Carcinogenesis 33 (2012) 427–435.
[112] A.E. Jaffe, R.A. Irizarry, Accounting for cellular heterogeneity is critical in epigenome-wide association studies, Genome Biol. 15 (2014).
[113] R. Lowe, V.K. Rakyan, Correcting for cell-type composition bias in epigenome-wide association studies, Genome Med. 6 (2014).
[114] M.J. Jones, S.A. Islam, R.D. Edgar, M.S. Kobor, Adjusting for cell type composition in DNA methylation data using a regression-based approach, in: P. Haggarty, K. Harrison (Eds.), Methods Mol. Biol, Humana Press, New York, NY, 2017, pp. 99–106.
[115] J. Brägelmann, J. Lorenzo Bermejo, A comparative analysis of cell-type adjustment methods for epigenome-wide association studies based on simulated and real data sets, Brief. Bioinform. 20 (2019) 2055–2065.
[116] S.A. Smallwood, H.J. Lee, C. Angermueller, F. Krueger, H. Saadeh, et al., Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity, Nat. Methods 11 (2014) 817–820.
[117] J. Ahn, S. Heo, J. Lee, D. Bang, Introduction to single-cell dna methylation profiling methods, Biomolecules 11 (2021).
[118] E.M. Martin, R.C. Fry, Environmental influences on the epigenome: exposure-associated DNA methylation in human populations, Annu. Rev. Public Health 39 (2018) 309–333.
[119] M. De Rosa, F. Azzato, J.E. Toblli, G. De Rosa, F. Fuentes, et al., A prospective observational cohort study highlights kidney biopsy findings of lupus nephritis patients in remission who flare following withdrawal of maintenance therapy, Kidney Int. 94 (2018) 788–794.
[120] M. Postal, J.F. Vivaldo, R. Fernandez-Ruiz, J.L. Paredes, S. Appenzeller, et al., Type I interferon in the pathogenesis of systemic lupus erythematosus, Curr. Opin. Immunol. 67 (2020) 87–94.
[121] T.B. Niewold, Connective tissue diseases: targeting type i interferon in systemic lupus erythematosus, Nat. Rev. Rheumatol. 12 (2016) 377–378.

Genome Medicine

RESEARCH

Open Access

Epigenetic and transcriptomic reprogramming in monocytes of severe COVID-19 patients reflects alterations in myeloid differentiation and the influence of inflammatory cytokines

Gerard Godoy-Tena[1], Anis Barmada[2,3], Octavio Morante-Palacios[1], Carlos de la Calle-Fabregat[1], Ricardo Martins-Ferreira[1], Anna G. Ferreté-Bonastre[1], Laura Ciudad[1], Adolfo Ruiz-Sanmartín[4], Mónica Martínez-Gallo[5], Ricard Ferrer[4], Juan Carlos Ruiz-Rodriguez[4], Javier Rodríguez-Ubreva[1], Roser Vento-Tormo[2] and Esteban Ballestar[1,6*]

Abstract

Background: COVID-19 manifests with a wide spectrum of clinical phenotypes, ranging from asymptomatic and mild to severe and critical. Severe and critical COVID-19 patients are characterized by marked changes in the myeloid compartment, especially monocytes. However, little is known about the epigenetic alterations that occur in these cells during hyperinflammatory responses in severe COVID-19 patients.

Methods: In this study, we obtained the DNA methylome and transcriptome of peripheral blood monocytes from severe COVID-19 patients. DNA samples extracted from CD14 + CD15- monocytes of 48 severe COVID-19 patients and 11 healthy controls were hybridized on MethylationEPIC BeadChip arrays. In parallel, single-cell transcriptomics of 10 severe COVID-19 patients were generated. CellPhoneDB was used to infer changes in the crosstalk between monocytes and other immune cell types.

Results: We observed DNA methylation changes in CpG sites associated with interferon-related genes and genes associated with antigen presentation, concordant with gene expression changes. These changes significantly overlapped with those occurring in bacterial sepsis, although specific DNA methylation alterations in genes specific to viral infection were also identified. We also found these alterations to comprise some of the DNA methylation changes occurring during myeloid differentiation and under the influence of inflammatory cytokines. A progression of DNA methylation alterations in relation to the Sequential Organ Failure Assessment (SOFA) score was found to be related to interferon-related genes and T-helper 1 cell cytokine production. CellPhoneDB analysis of the single-cell transcriptomes of other immune cell types suggested the existence of altered crosstalk between monocytes and other cell types like NK cells and regulatory T cells.

*Correspondence: eballestar@carrerasresearch.org

[1] Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain
Full list of author information is available at the end of the article

Conclusion: Our findings show the occurrence of an epigenetic and transcriptional reprogramming of peripheral blood monocytes, which could be associated with the release of aberrant immature monocytes, increased systemic levels of pro-inflammatory cytokines, and changes in immune cell crosstalk in these patients.

Keywords: COVID-19, Monocytes, Epigenomics, DNA methylation, Single-cell transcriptomics, Immune cell crosstalk

Background

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes the well-known Coronavirus disease 2019 (COVID-19), which has become a major global health burden. SARS-CoV-2 infection occurs through the nasopharyngeal mucosa [1]. Subsequent immune responses occur at the local mucosa and at a systemic level. An effective response to SARS-CoV-2 infection requires coordination between the innate and adaptive immune systems, including granulocytes, monocytes, macrophages, and T and B cells [2, 3]. The range of immune responses to SARS-CoV-2 infection is diverse, from asymptomatic or mild upper-respiratory illness to severe viral pneumonia, acute respiratory distress syndrome, and death [4]. The most severe forms of COVID-19 are caused by dysregulation of immune homeostasis, which leads to hyperinflammation in the lungs [5]. This has been shown to be more pronounced in the elderly and in individuals with pre-existing comorbidities [6, 7]. Nevertheless, despite the numerous studies performed in the field, the impact of exacerbated immune responses associated with severe COVID-19 at the systemic level remains unclear.

Various studies have demonstrated that peripheral pathogenic T cells and inflammatory monocytes can induce a cytokine storm in severe COVID-19 patients [8]. This takes the form of excessive production of inflammatory mediators, specifically, interleukin (IL)-6, IL-1β, granulocyte–macrophage colony-stimulating factor (GM-CSF), tumor necrosis factor-alpha (TNFα), and interferon gamma (IFNγ) [8–11]. IFN is essential for inducing the innate immune response during viral infection through different interferon regulatory factors (IRFs) [12]. Further, in COVID-19 patients, type I IFN deficiency appears to be a hallmark of severe cases [13–19] in association with persistent blood viral load and an exacerbated inflammatory response [14].

Single-cell omics studies have identified specific transcriptional features in monocytes, natural killer (NK) cells, dendritic cells (DCs), and T cells associated with the severity of COVID-19 [13, 20–22]. These studies have revealed that severe COVID-19 is marked by a dysregulated myeloid cell compartment [13]. It has also been shown that monocytes from severe COVID-19 patients are characterized by a tolerogenic phenotype with reduced expression of class II major histocompatibility complex (MHC-II) antigens [23] and increased activation of apoptotic pathways [24].

Differentiation and activation of monocytes and other myeloid cells are directly associated with epigenetic mechanisms [25]. The functional plasticity of these cells is also reflected at the epigenetic level, and several studies have shown that DNA methylation profiles, among other epigenetic marks, vary in response to inflammatory cytokines, hormones, and other factors [26, 27], depending on their functionality. Cytosine methylation (5mC) occurs at CpG dinucleotides and is generally associated with transcriptional repression [28], although its relationship with transcription depends on the genomic location of the affected CpG sites [29]. In some cases, DNA methylation changes occur as a result of upstream environmental effects that link cell membrane receptors, signaling pathways, and transcription factors (TFs) that can either directly recruit DNA methyltransferases (DNMT) and ten–eleven translocation (TET) enzymes, or indirectly influence their binding to specific genomic sites.

The characterization of the epigenetic and transcriptomic reprogramming in monocytes, given their central role in inflammatory responses, is essential if we are to understand the specific dysregulated pathways involved in severe forms of COVID-19. In this study, we obtained the DNA methylation profiles of peripheral blood monocytes of severe COVID-19 patients and studied their relationship with transcriptomic changes, obtained by generating droplet-based single-cell RNA sequencing (scRNA-seq) data from peripheral blood.

Methods

Human samples

Our study included a selection of 58 severe COVID-19 patients from the intensive care unit (ICU) of Vall d'Hebron University Hospital (Barcelona) recruited during the second wave of infection in Spain (October to November 2020). Peripheral blood samples were taken at different times following admission of the patient to the ICU, as specified in Additional file 1. Table S1 (Days in ICU). Ninety-four percent of the patients required intubation and all enrolled cases were confirmed to be infected with SARS-CoV-2 using real-time RT-PCR at the time of collection. For all enrolled patients, the date of enrollment, clinical classification, or treatment

Table 1 Summary of patient cohort for DNA methylation analysis

	Healthy controls	COVID19 severe patients	*p* value
Number	11	48	-
Age (mean ± SD)	50 ± 11.16	60 ± 11.96	0.0042
Sex (% female)	36.4	25	0.710
SOFA	0	5 ± 2.97	2.4e − 07
IL-6 (pg/ml) (mean ± SD)	NA	316.94 ± 1238.82	-
Days in ICU (mean ± SD)	NA	6 ± 5.93	-
Treated with dexamethasone (%)	NA	52.08	-
Obesity (%)	NA	27.03	-
Hypertension (%)	NA	56.25	-
Death (%)	NA	33.33	-
Mechanical ventilation (%)	NA	93.75	-

was obtained from the clinical records. From all these patients, 48 of the 58 patients were selected for DNA methylation analysis (Additional file 1. Table S1) and peripheral blood mononuclear cells (PBMCs) from 10 of the 58 patients were used for droplet-based scRNA-seq analysis (Additional file 2. Table S2). The control population for the DNA methylation analysis comprised 11 healthy donors (HDs) recruited at the Blood Bank of Vall d'Hebron University Hospital. Table 1 summarizes the characteristics and clinical data from patients included in the DNA methylation analysis. We included an additional group of 14 patients from the same hospital for DNA methylation and expression validation, including 9 severe COVID-19 patients and 5 mild COVID-19 patients, together with an additional group of 6 HDs. The validation cohort was collected during February 2022, applying the same selection criteria as for the discovery cohort. For the validation cohort, we only included non-vaccinated patients, to match the vaccination status with that of the patients collected in the initial phase of the study. Clinical information corresponding to the new cohort is also included in Additional file 1. Table S1 (validation cohort). This study was approved by the Clinical Research Ethics Committees of Hospital Universitari Germans Trias i Pujol (PI-20–129) and Vall d'Hebron University Hospital (PR(AG)282/2020), both of which adhered to the principles set out in the WMA Declaration of Helsinki. Informed consent was obtained from all patients before their inclusion.

Monocyte purification and DNA isolation

PBMCs were obtained from peripheral blood by Ficoll gradient using Lymphocyte Isolation Solution (Rafer, Zaragoza, Spain) from 48 of the severe COVID-19 patients and 11 HDs. Once PBMCs were isolated, all samples were stored at − 150 °C in 10% DMSO in fetal bovine serum (FBS) until monocyte purification. The monocyte population was isolated by flow cytometry (FacsAria Fusion, BD, Beckton Dickinson, San Jose, CA, USA). PBMCs were stained with CD14-APC-Vio770 (Miltenyi Biotec) and CD15-FITC (Miltenyi Biotec) in staining buffer (MACS) for 20 min. A gating strategy was employed to eliminate cell debris, doublets, and DAPI + cells. CD14 and CD15 antibodies were used to isolate CD14 + CD15 − . Purified cells were pelleted and stored at − 80 °C.

After monocyte isolation, DNA was isolated using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) following the manufacturer's instructions.

DNA methylation profiling

Bisulfite (BS) conversion was performed using EZ-96 DNA Methylation™ Kit (Zymo Research, CA, USA) according to the manufacturer's instructions. Five hundred nanograms of BS-converted DNA was hybridized on Infinium Methylation EPIC BeadChip arrays (Illumina, Inc., San Diego, CA, USA). These were used to analyze DNA methylation. They enable > 850,000 methylation sites per sample to be assessed at single-nucleotide resolution, which corresponds to 99% of the reference sequence (RefSeq) genes.

Each methylation data point was obtained from a combination of the Cy3 and Cy5 fluorescent intensities from the methylated and unmethylated alleles. Background intensity computed from a set of negative controls was subtracted from each data point. For representation and further analysis, we used *beta* (*b*) and *M* values. *Beta* is the ratio of methylated probe intensity to overall intensity (the sum of the methylated and unmethylated probe intensities). *M* is calculated as the $\log_2$ ratio of the intensities of the methylated and unmethylated probes. For statistical purposes, the use of *M* is more appropriate

since *b*-values are severely heteroscedastic for highly methylated and unmethylated CpG sites. Raw DNA methylation data are available at GEO, with accession number GSE188573 [30].

Quality control, data normalization, and statistical analysis of DMPs

Quality control and analysis of EPIC arrays were performed using ShinyÉPICo [31], a graphical pipeline that uses *minfi* (v1.36) [32] for normalization, and *limma* (v3.46) [33] for analyzing differentially methylated positions. ShinyÉPICo is available as an R package at the Bioconductor (http://bioconductor.org/packages/shinyepico/) and GitHub (https://github.com/omorante/shinyepico) sites. We used the BS conversion control probe test included in ShinyÉPICo to determine whether the conversion rate was above the quality threshold of 2 established by Illumina. The threshold was calculated from the information of the BS conversion control probes of the EPIC arrays. When the BS conversion reaction is successful, control probes display strong signal in the red channel, whereas if the sample has unconverted DNA, control probes have a strong signal in the green channel. The red/green ratio for each control position was calculated for each sample.

CpH and SNP loci were removed by the Noob method, followed by quantile normalization. Sex chromosomes (X and Y) were also excluded from the analysis to avoid discordant information among samples. Even when data were generated in a single batch and randomized, we applied the batch effect correction. Sex and age of the donors were included as covariates, to minimize confounding effects due to differences between the median age of the patient and control cohorts, and the Trend and Robust options were implemented in the eBayes moderated *t*-test analysis. To compare healthy donors with the entire severe COVID-19 patient cohort, we identified differentially methylated CpG sites by using *t*-tests and a method with defined empirical array weights, included in the limma package [33], and selecting CpGs with a false discovery rate (FDR) of <0.05 and a $\Delta\beta$ of >0.15. To test the effects of potential changes in monocyte subset proportions, we also included this information as a covariate, and performed the same analysis as above, but including only those samples for which such information was available.

We used the iEVORA package (v1.9.1) [34] to identify differentially variable positions (DVPs). This algorithm identifies differences in variance using Bartlett's test (FDR <0.001), followed by the comparison of means using *t*-test ($p<0.05$) to regularize the variability test, which is overly sensitive to single outliers. For the analysis in Fig. 2, we calculated Spearman's correlation coefficient (rho) to measure the association of two variables and thereby identify CpG sites in which DNA methylation was correlated with SOFA in patients with severe COVID-19. We selected the CpG sites for which Spearman's rho was greater than 0.4 and had an associated value of $p<0.01$. Principal component analysis (PCA) of *b*-values from ShinyÉPICo was used to determine the correlations of PCs with clinical variables such as dexamethasone treatment, obesity, and hypertension. Pearson correlation coefficients between numerical variables and PCs were calculated. Categorical variables were entered in a linear model together with the PCs, which were considered as a function of the variable.

Gene ontology, transcription factor enrichment, and chromatin state discovery and characterization

The GREAT (v3.0.0) online tool (http://great.stanford.edu/public/html) was used for gene ontology (GO) analysis, in which genomic regions were annotated by applying adapted basal and extension settings (5 kb upstream, 5 kb downstream, 1000 kb plus distal). GRCh37 (UCSC hg19, Feb. 2009) was used as the alignment genome reference. Annotated CpGs in the EPIC array were used as background. GO terms were considered significant for a $>$ twofold change and an FDR <0.05. Enrichment is represented as $-\log_2$(FDR). GO categories with $p<0.05$ were considered significantly enriched. GO analysis of differentially expressed genes (DEGs) was carried out using the online Enricher gene ontology analysis tool (https://maayanlab.cloud/Enrichr/). GO categories with a $>$ twofold change and an FDR <0.05 were considered significantly enriched.

We used the *findMotifsGenome.pt* tool from the motif discovery HOMER software (v4.10.3) to analyze motif enrichment [35]. A flanking window of ± 250 bp from each CpG was applied, and CpGs annotated in the EPIC array were used as background. To determine the location relative to a CpG island (CGI), we used "hg19_cpgs" annotation in the *annotatr* (v1.8) R package. The statistical test used for the enrichment in these analyses was Fisher's exact test. Chromatin functional state enrichment of DMPs was measured using public PBMC data from the Roadmap Epigenomics Project (http://www.roadmapepigenomics.org/) generated with ChromHMM (v1.23) [36]. Enrichments were calculated with Fisher's exact test using array annotation as background regions. Only significantly enriched states are shown.

Heatmaps and PCA plots

Heatmaps of DMPs were generated with functions available in the *ComplexHeatmap* (v2.11.1) and *gplots* (v3.1.3) R packages. We used PCA for the low-dimensional analyses. PCA projection matrices were calculated with R's

prcomp function, and visual representations of PCs were plotted with the *ggfortify* package (v4.1.4).

Whole-genome bisulfite sequencing (WGBS) analysis

DNA methylation values of *Ensembl Regulatory Build* regions of progenitor cells such as hematopoietic stem cell (HSC), multipotent progenitor (MPP), common myeloid progenitor (CMP), granulocyte macrophage progenitor (GMP), and control monocytes were extracted from public whole-genome bisulfite sequencing (WGBS) (GSE87197) [37]. Using GenomicRanges (v1.42.0) and based on genomic location, the overlap of the hypermethylated DMPs observed in COVID-19 compared with HD was determined with the *Ensembl Regulatory Regions* from the hematopoietic precursors and monocytes. For this analysis, all DNA methylation data were annotated with respect to the GRCh38 human genome reference.

Single-cell capture

PBMCs from 10 ICU patients were used to generate single-cell gel beads-in-emulsion (GEMs) (Additional file 2. Table S2). Cells were then washed three times and counted. For samples with low viability (< 90%), we performed Ficoll separation in an Eppendorf tube to eliminate dead cells and increase cell viability. For samples with greater than 90% viability, we filtered using a Flowmi strainer and counted the cells before loading into 10X chromium to generate single-cell GEMs, following the manufacturer's instructions. We loaded 50,000 cells per pool, including a total of 4 patients per pool. Datasets from patients and HDs are available as h5ad files (https://www.COVID-19cellatlas.org/index.patient.html (Additional file 2. Table S2). In parallel, genomic DNA was isolated from the same 10 PBMCs for genotyping and subsequent donor deconvolution (as described in [38]) using a Maxwell® 16 Blood DNA Purification Kit from Promega following the manufacturer's instructions.

scRNA-seq cell type identification and annotation

Single-cell transcriptome data from COVID-19 patients were quantified and aligned using Cell Ranger (v3.1) with the GRCh38 genome concatenated to SARS-Cov-2 genome as a reference. Thereafter, cells from pooled samples were deconvolved and demultiplexed using *Souporcell* (v3.0) [39], yielding a genotype variant that allows donor identity to be matched across different samples. This additionally enabled the removal of doublet cells that could not be explained by any single genotype. *Scrublet* (v0.2.3) [40] was subsequently employed to further filter out other doublets based on computed doublet scores. Specifically, Student's *t*-test ($p < 0.01$) after Bonferroni correction was used within fine-grained sub-clustering of each initial cluster produced by the Leiden algorithm. Data were not denoised because no significant contamination or ambient RNA was present. Previously described scRNA-seq datasets of HDs [41] were then integrated for comparison using single-cell variational inference (scVI) [42] with a generative model of 64 latent variables and 500 iterations. More specifically, scVI employs a negative binomial model using raw counts, selecting 5000 highly variable genes to produce the latent variables. Defined cell-cycle phase-specific genes in the *Seurat* package (v4.1.0) [43] were excluded from these to reduce the dependence of clustering on cell-cycle effects. Data were subsequently analyzed using *Scanpy* (v1.9.1) [44] following the recommended standard practices. For quality control, genes expressed in fewer than three cells, and cells with fewer than 200 genes or more than 20% mitochondrial gene content, were removed prior to downstream analysis. Data were normalized (scanpy.pp.normalize_per_cell, scaling factor = 10,000) and $\log_2$-transformed (scanpy.pp.log1p). For gene expression visualization (e.g., heat maps), data were further scaled (scanpy.pp.scale, maximum value = 10).

Cell type clustering and annotation

The resulting latent representation from the integrated datasets was used to compute the neighborhood graph (scanpy.pp.neighbors), then the Louvain clustering algorithm (scanpy.tl.louvain, resolution = 3) and Uniform Manifold Approximation and Projection (UMAP) visualization (scanpy.tl.umap) were employed. Cell type annotations were manually refined using literature-driven, cell-specific marker genes. Identified residual RBCs from incomplete PBMC isolation were excluded before further analysis, as recommended [45].

Differential gene expression and transcription factor-enrichment analysis

Differential gene expression between COVID-19 patients and healthy individuals (FDR < 0.05) was analyzed using the *limma* package [46]. To predict transcription factor (TF) involvement in transcriptomic changes, we used DoRothEA (Discriminant Regulon Expression Analysis) v2 tool [47]. Regulons with a confidence score of A–C were analyzed, and cases with $p < 0.05$ and a normalized enrichment score (NES) of ± 2 were considered significantly enriched.

Cell–cell communication

Based on the differential expression analysis, CellPhoneDB [48] v3 (www.CellPhoneDB.org) was used to infer changes in ligand/receptor interactions between the identified cell types in COVID-19 versus HD. Specifically, instead of random shuffling, as used in the previously

described statistical method, differentially expressed genes (FDR < 0.05) were used to select interactions that were significantly enriched in either severe COVID-19 patients or healthy individuals relative to the other group. An interaction was considered enriched if at least one of the two partners (ligand or receptor) was differentially expressed, and if both partners were expressed by at least 10% of the interacting cells.

Bisulfite pyrosequencing

EZ DNA Methylation-Gold kit (Zymo Research) was used to BS-converted 500 ng of genomic DNA following the manufacturer's instructions. BS-treated DNA was PCR-amplified using IMMOLASE DNA polymerase kit (Bioline). Primers used for the PCR were designed with PyroMark Assay Design 2.0 software (Qiagen) (Additional file 3. Table S3). PCR amplicons were pyrosequenced with the PyroMark Q24 system and analyzed with PyroMark Q48 Autoprep (Qiagen).

Real-time quantitative polymerase chain reaction (RT-qPCR)

The Transcriptor First Strand cDNA Synthesis Kit (Roche) was used to convert 250 ng of total RNA to cDNA following the manufacturer's instructions. RT-qPCR primers were designed with Primer3 software [49] (Additional file 3. Table S3). RT-qPCR reactions were prepared with LightCycler 480 SYBR Green I Master (Roche) according to the manufacturer's instructions and analyzed with a LightCycler 480 instrument (Roche).

Flow cytometry

To study the surface cell markers on monocytes (CD14+), PBMCs from the 10 patients used for single-cell analysis and 10 HDs were defrosted and washed once with PBS. After blocking for non-specific binding with Fc block (BD Pharmingen) for 5 min on ice, cells were incubated for 20 min on ice using staining buffer (PBS with 4% fetal bovine serum and 0.4% EDTA). Antibodies used included the following: CD14-FitC (Miltenyi Biotec), CD85-PEvio770 (Miltenyi Biotec), CD172a-APC (Miltenyi Biotec), CD97-PEvio770 (Miltenyi Biotec), CD31-PE (Miltenyi Biotec), CD366-PEvio615 (Miltenyi Biotec), CD62L-APC (Miltenyi Biotec), CD58-PE (Miltenyi Biotec), CD191-PEvio770 (Miltenyi Biotec), CD52-PEvio615 (Miltenyi Biotec), CD48-APC (Miltenyi Biotec). Cells were analyzed in a BD FACSCanto-II flow cytometer.

Statistical analysis

All statistical analyses were done with R v4.0.2. Box, bar, violin, bubble, and line plots were generated using functions from the *ggplot2* (v3.3.6) and *ggpubr* (v4.0) packages. Mean normalized DNA methylation values were compared using two-tailed test. Multivariate frequency distributions were calculated using Fisher's exact test. The levels of significance are indicated as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

Results

DNA methylome remodeling in peripheral blood monocytes of severe COVID-19 patients

To directly inspect epigenetic alterations in peripheral blood monocytes in severe COVID-19, we isolated CD14+CD15− cells from 59 blood samples, comprising 48 severe COVID-19 patients and 11 healthy donors (HDs), and performed DNA methylation profiling (Fig. 1A, Table 1, and Additional file 1. Table S1). For cell sorting, we first separated live cells from debris, then extracted singlets and isolated CD14+CD15− cells to avoid neutrophil contamination (Fig. 1B) [50]. Since we selected CD14+ cells, the purification procedure only included classical (CM) (CD14+CD16−) and intermediate monocytes (IM) (CD14+CD16+), excluding the non-classical monocyte (NCM) (CD14lowCD16+) subpopulation, which in healthy individuals corresponds to around 5% of the total monocyte compartment [51]. Negative selection using CD15 was necessary, as there is a significant increase in the frequency of neutrophils in severe COVID-19 patients, as activated neutrophils are not separated in the Ficoll step [52] (Additional file 4. Figure S1A-S1C). To confirm the purity of our monocytes, we performed FACS analysis and obtained an average purity of 98% (example in Additional file 4 Figure S1D). Studies in various other inflammatory diseases

(See figure on next page.)

Fig. 1 Analysis of DNA methylation in blood monocytes of severe COVID-19 patients. **A** Scheme depicting the cohort and workflow for monocyte purification of severe COVID-19 patients and controls and DNA methylation analysis. **B** Representative flow cytometry profile, indicating sorting gates used to purify monocytes from HD and COVID-19 patients' peripheral blood. **C** Scaled DNA methylation (*z*-score) heatmap of differentially methylated positions (DMPs) between HDs (blue bar above) and COVID-19 patients (red bar above). Significant DMPs were obtained by applying a filter of FDR > 0.05 and a differential of beta value (Δß) > 0.15. A scale is shown on the right, in which blue and red indicate lower and higher levels of methylation, respectively. Clinical and treatment data of COVID-19 patients are represented above the heatmap. SOFA, IL-6 level, and days in the ICU scales are shown on the right of the panel **D** Principal component analysis (PCA) of the DMPs. HDs and severe COVID-19 patients are illustrated as blue and red dots, respectively. **E** Gene ontology of hypermethylated and hypomethylated DMPs. Selected significant functional categories (FDR < 0.05) are shown. **F** Bubble plot of TF motifs enriched on hypermethylated and hypomethylated DMPs. Bubbles are colored according to their TF family; their size corresponds to the FDR rank. **G** Box plot of individual DNA methylation values of CpG from hypermethylated and hypomethylated clusters (*b*-values), with the name of the closest gene and the position relative to the transcription start site

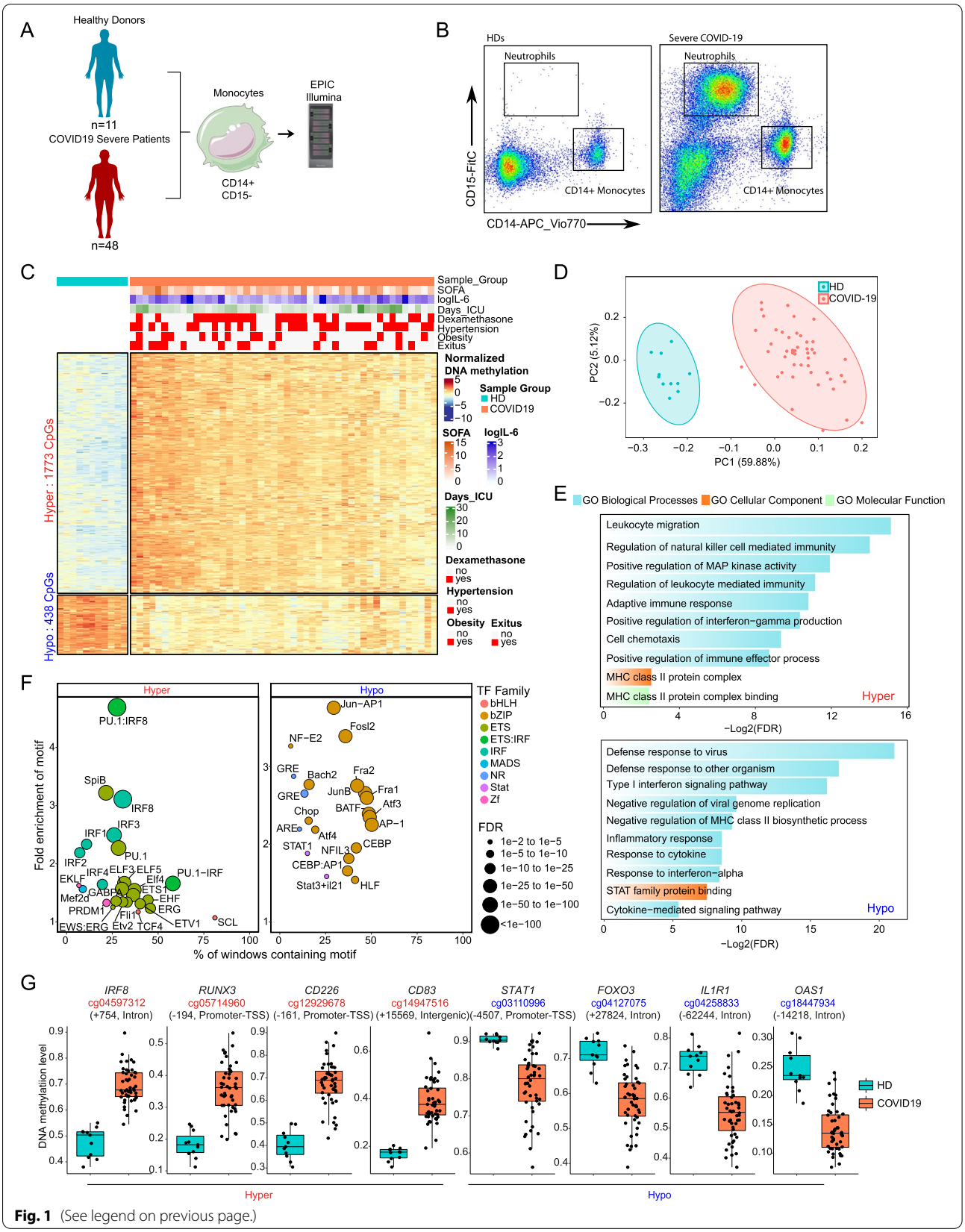


Fig. 1 (See legend on previous page.)

have shown that the proportions of monocytes can shift between the three major subsets, i.e., CM, IM, and NCM. For instance, it has been shown that severe COVID-19 patients feature reduced NCM and IM populations [53]. The analysis of monocyte subpopulations in our cohort showed a significant increase in the CM population and a decrease in the NCM population (Additional file 4. Figure S1E-S1F). Since we purified CD14 + monocytes, our study only included CM and IM.

We performed DNA methylation profiling of isolated monocytes and identified 2211 differentially methylated positions (DMPs) of CpGs in severe COVID-19 patients compared with HDs (FDR < 0.05 and absolute Δß > 0.15). Of these, 1773 were hypermethylated (hypermethylated cluster) and 438 were hypomethylated (hypomethylated cluster) (Fig. 1C and Additional file 5. Table S4). PCA of these DMPs showed that the two groups of monocytes (COVID-19 and HD) separated along the first principal component axis (Fig. 1D). We obtained similar results when we included monocyte subpopulation proportions as a covariate in the analysis (overlap, $p < 0.0001$) (Additional file 6. Figure S2A). No significant differences (FDR < 0.05) were observed within COVID-19 patients separated by their condition (obesity, hypertension, days admitted to the ICU, and exitus/death) or treatment with dexamethasone (Additional file 1. Table S1). None of the abovementioned conditions was significantly correlated with the DNA methylation changes (Additional file 6. Figure S2B). This was also apparent from the PCA showing the overlap of patients with different clinical parameters (Additional file 6. Figure S2C).

The analysis of the genomic functional features of the DMPs in the hypermethylated and hypomethylated clusters (Additional file 6. Figure S2D) using public data from monocytes [36] revealed an enrichment in promoters and enhancers. This is consistent with their proposed roles for DNA methylation in regulatory elements [54].

Gene ontology analysis (GO) of the two DMP clusters revealed several functional categories associated with the immune response to viral infection (Fig. 1E). In the hypermethylated cluster, we observed enrichment of categories such as natural killer-mediated immunity, leukocyte migration, adaptive immune response, and positive regulation of interferon gamma production. We also observed hypermethylation in the MHC-II protein complex that was related to antigen presentation. In addition, we found an enrichment of the positive regulation of MAP kinase activity category (Fig. 1E, top panel). In the hypomethylated cluster, we observed enrichment of functional categories relevant to viral infection, including defense response to virus and negative regulation of viral genome replication. Importantly, the hypomethylated cluster also featured enrichment of functional categories related to type I interferons (IFN) signaling and MHC class II (Fig. 1E, bottom panel).

Transcription factor (TF) binding motif enrichment analysis, in 250-bp windows surrounding DMPs, revealed overrepresentation of TFs of significance to the immune response. The hypermethylated cluster CpGs displayed enrichment of binding motifs of IRFs and ETS TF families, which are linked to IFN changes (Fig. 1F, left panel). Motifs of the bZIP TF family like AP-1, Jun, Fosl2, Fra1, and Fra2 were enriched in the hypomethylated cluster. DMPs of the hypomethylated cluster were also enriched in motifs of the signal transducer factor and activator of transcription factor (STAT) members STAT1 and STAT3. We also detected enrichment of the glucocorticoid response element (GRE) in the hypomethylated cluster (Fig. 1F, right panel). Given these results, we hypothesized that pharmacological treatment with glucocorticoids (GCs) in severe COVID-19 patients in the intensive care unit (ICU) might influence DNA methylation in monocytes. To test this possibility, we performed *limma* analysis and subsequent binding motif enrichment after separating COVID-19 patients into two groups, with and without GC treatment. Both groups of patients exhibited significant enrichment of GRE motifs in the hypomethylated cluster (Additional file 6. Figure S2E), suggesting that the endogenous production of GCs in severe COVID-19 patients could participate in the hypomethylation through GRE. However, given the size of the cohort, we cannot rule out the possibility that pharmacological treatment could also influence DNA methylation changes and therefore remains as a potential confounder factor.

Inspection of the individual genes within or in the vicinity of the DMPs revealed several genes with functions essential to the viral immune response. The list of relevant genes included *IRF8*, *RUNX3, CD226*, and *CD83* in the hypermethylated cluster, and *STAT1*, *FOXO3*, *IL1R1,* and *OAS1* in the hypomethylated cluster (Fig. 1G). We validated these results using bisulfite pyrosequencing in a new cohort of severe COVID-19 patients (Additional file 6. Figure S2F). Interestingly, these changes were also observed in mild COVID-19 patients (Additional file 6. Figure S2F). *IRF8*, *IL1R1,* and *CD83* are associated with the IFN response. *CD226* encodes a glycoprotein related to monocyte, NK, and T cell adhesion. This glycoprotein has been shown to be involved in the cytotoxicity of these cells and is known to be altered in COVID-19 patients [13]. *STAT1* is associated with the cytokine response, which, in turn, is related to *IL1R1*. The latter is the receptor of interleukin 1, which participates in the inflammatory response and is strongly expressed in severe COVID-19 patients [14]. *OAS1* is induced by interferons and activates latent RNase, causing viral RNA

degradation, which could be related to the identification of the category negative regulation of viral genome replication in the GO analysis.

Monocytes from severe COVID-19 patients display increased DNA methylation variability

Overall, our DNA methylation analysis showed greater heterogeneity (different variable positions, DVPs) in the profiles from COVID-19 patient monocytes than in those from HDs (Additional file 6. Figure S2G). We then examined the relationship between the DNA methylation profiles and the Sequential Organ Failure Assessment (SOFA) score, which is used in ICUs to calculate organ damage. The score ranges from 0 to 24, with values greater than 6 being associated with a significant increase in the risk of mortality [55]. Using Spearman's correlation coefficient to assess specific hypermethylated or hypomethylated CpGs with SOFA, we identified 1375 CpG sites whose methylation levels positively correlated with SOFA (increased methylation) (rho < 0.4 and $p < 0.01$) and 1497 CpG sites with an inverse correlation with SOFA (decreased methylation) (rho < −0.4 and $p < 0.01$) (Fig. 2A and Additional file 7. Table S5). The mean normalization DNA methylation profiles of increased and decreased methylation CpG sites were similar in patients with low SOFA (< 6) and in healthy controls in an unsupervised representation but differed between the low and high SOFA score groups (Fig. 2B). These results suggest that changes in DNA methylation are concomitantly exacerbated for higher SOFA scores, which is associated with bad prognosis. Several CpGs correlating with SOFA were associated with genes, such as *IL17R*, *SOCS5,* and *PCDHA5*, that are involved in T cell-mediated inflammatory responses (Fig. 2C). Others, like *FOXG1* and *CDC20B*, are associated with DNA damage. GO analysis revealed that changes in DNA methylation that are concomitant with SOFA show an overrepresentation of terms associated with IFNγ, production of the molecular mediator involved in inflammatory response, viral gene expression, the B cell proliferation involved in immune response, and Th1 cell cytokine production (Fig. 2D).

DNA methylation alterations in monocytes of severe COVID-19 patients significantly associate with those derived from patients with bacterial sepsis, myeloid differentiation, and the influence of inflammatory cytokines

To better characterize the impact of DNA methylation changes in COVID-19, we compared the DMPs from severe COVID-19 patients with those obtained from monocytes derived from patients with bacterial sepsis in a previous study by our team [27], given that severe COVID-19 can be considered a form of sepsis [56]. To this end, we first estimated the DNA methylation values of DMPs corresponding to the sepsis relative to the HD comparison from our previous sepsis study (accession number GSE138074) [27] using the data from the severe COVID-19 methylation dataset. Overall, we found significant enrichment in the hypermethylation and hypomethylation clusters (Fig. 3A). We also calculated the odds ratio of the overlap between these two datasets and found a strong enrichment of the hyper-DMPs in COVID-19 relative to those in sepsis (FDR ≤ $2.22 \cdot 10^{-16}$) and in the hypo-DMPs (FDR ≤ $2.22 \cdot 10^{-16}$) (Fig. 3B). We also confirmed an enrichment in introns and depletion in promoters relative to the background when testing the genomic location of the DMPs common to both COVID-19 and sepsis (Fig. 3C and Figure S3A). DMPs located in introns are often localized in enhancer regions involved in long-distance regulation [54].

We then determined that the two datasets had 362 hypermethylated and 92 hypomethylated CpGs in common (Fig. 3D), corresponding to 51% of the total DMPs of the sepsis patients (Additional file 8. Figure S3B). GO analysis of the shared DMPs revealed significant enrichment in functional terms related to host response, including regulation of NK cells, inflammatory response, and leukocyte chemotaxis (Additional file 8. Figure S3C). Shared hypermethylated CpGs were enriched in functional categories related to cell signaling, such as the JAK-STAT and MAPK pathways, that could be involved in the reduction of the inflammatory response and the IL15- and IL12-mediated signaling pathways, which are related to cytokine production and Th1 proliferation (Fig. 3E, left panel). Shared hypomethylated CpGs were enriched in functional categories responsible for regulating the inflammatory response, such as negative regulation of IL-1 production and positive regulation of macrophage activation. In concordance with the hypermethylated cluster, we also observed negative regulation of IFNα production (Fig. 3E, right panel). It is of note that severe COVID-19-specific DMPs were enriched in functional categories related to virus infection, such as the defense response to virus, and impairment of the antigen-presenting process, which seems to be specific to COVID-19 infection [13, 23] (Additional file 8. Figure S3D).

Inspection of TF binding motifs corresponding to the DMPs shared between the two groups, separating the shared hypermethylated and hypomethylated CpG sets revealed IRF family transcription factors like IRF1, IRF2, IFR3, and IRF8 in the shared hypermethylated CpG set, which are well established regulators of the type I IFN system, being common in viral and bacterial infections [57]. We also detected enrichment of the ETS transcription factors that are regulated by MAPK proteins, which were enriched in the GO analysis (Fig. 3F). In the shared

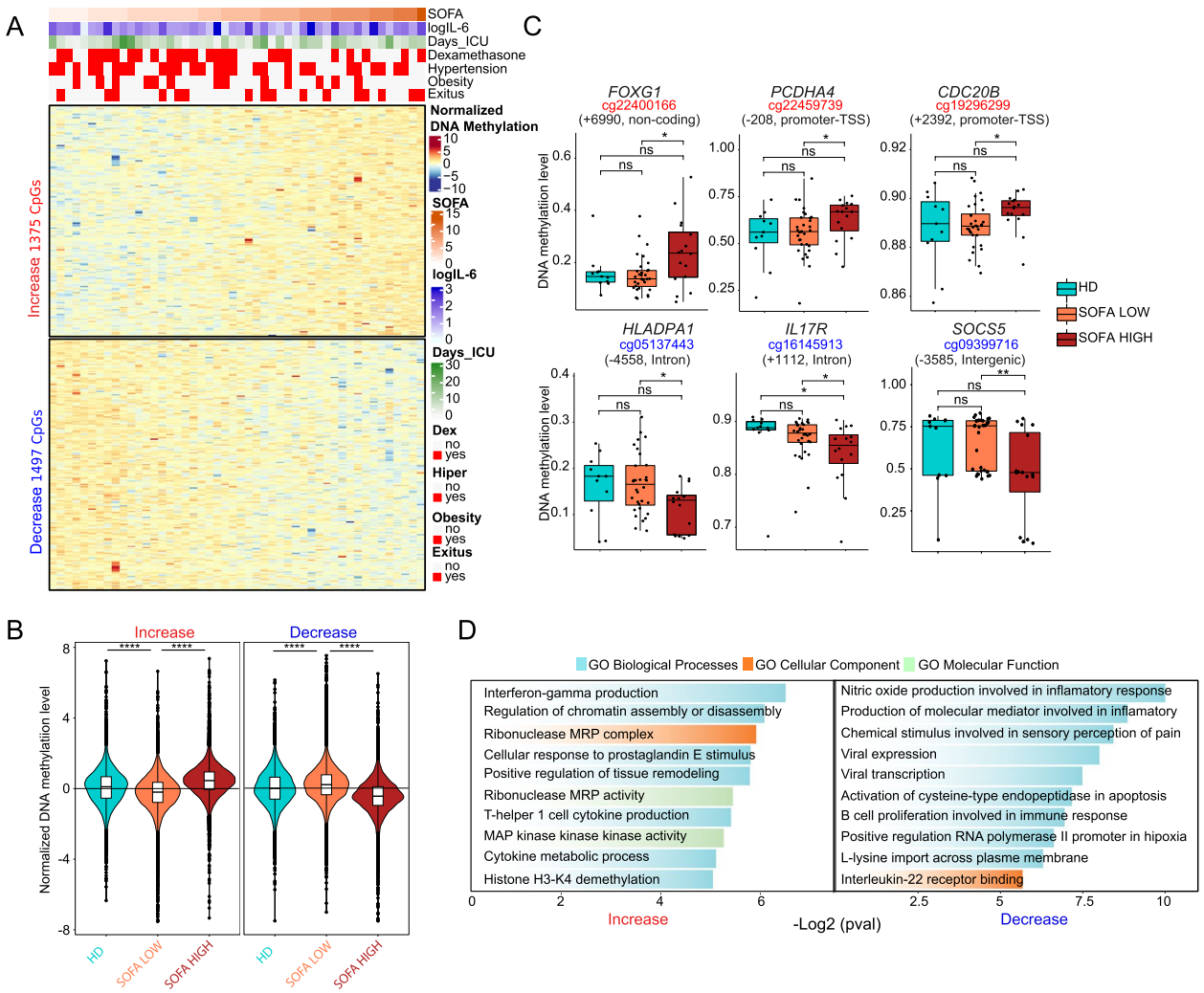


Fig. 2 DNA methylation changes in COVID-19 monocytes parallel organ damage. **A** Heatmap of severe COVID-19 patients with DNA methylation ordered by SOFA score, including all CpG-containing probes significantly correlated with the SOFA score (Spearman correlation coefficient rho > 0.4, $p < 0.01$). Clinical and treatment data of COVID-19 patients are shown above the heatmap. SOFA, IL-6 level, and days in the ICU scales are shown on the right of the panel **B**. Normalized methylation values from heatmap showing overall group methylation of HD. Patients with SOFA ≤ 6 are indicated as SOFA LOW; those with SOFA > 6 are indicated as SOFA HIGH. **C** DNA methylation levels (*b*-values) of selected individual CpGs (and closest genes) in hypermethylated and hypomethylated sets and their position relative to the transcription start site. **D** Gene ontology (GO) analysis of hypermethylated and hypomethylated DMPs, analyzed with the GREAT online tool, in which CpG annotation in the EPIC array was used as background. Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

hypomethylated set, we noted enrichment of STAT3 and TFs from bZIP AP-1, like Jun, and other bZIPs, like CEBP. Interestingly, GRE was also present in the shared hypomethylated cluster (Fig. 3F). This suggests the influence of GC in the acquisition of aberrant methylation profiles in COVID-19 and sepsis. Individual genes associated with the COVID-19/sepsis shared hypermethylated and hypomethylated CpG genes include type I IFN-related genes, like *IRF2*, and others, such as *IL1A* and *CCR2*, that are involved in inflammatory processes and monocyte chemotaxis, respectively (Fig. 3G). We also identified several genes among the shared hypomethylated set, like *CD163*, *SOCS1*, and *IL10*, that have been associated with the acquisition of tolerogenic properties in monocytes [58] (Fig. 3G).

In both infections, systemic inflammation could be responsible for part of the DNA methylation changes that arise in monocytes. To address this possibility, we examined the DNA methylation levels of the hypomethylated and hypermethylated CpGs of severe COVID-19 and sepsis patients in monocytes isolated from healthy donor PBMCs that had been treated in vitro with inflammatory

cytokines like IFNα, IFNγ, and TNFα [26] (accession number GSE134425). This analysis revealed several significant changes following the trends for both COVID-19 and sepsis (Additional file 8. Figure S3E), suggesting that these inflammatory cytokines, which are elevated in these patients, could influence the monocyte DNA methylomes.

An alternative explanation for the observed changes in severe COVID-19 monocyte methylomes could be that DNA methylation changes reflect alterations during myeloid/monocyte differentiation or the release of immature or aberrant monocytes. This has been described in severe COVID-19 cases [13, 59–62]. It is worth noting that immature cells are also released from the bone marrow in sepsis [63]. To test this hypothesis, we used public whole-genome bisulfite sequencing (WGBS) data (GSE87197) of progenitor cells including HSC, MPP, CMP, and GMP cells and monocytes as references. We compared the 1773 hypermethylated CpGs based on their genomic location and obtained 1511 unique Ensembl Regions, which grouped in two clusters. Cluster 1 showed low-level demethylation in monocytes compared with all hematopoietic precursor cell types, whereas cluster 2 showed clear demethylation in monocytes (Fig. 3H). These results are compatible with the possibility that a proportion of the DMPs in severe COVID-19 result from aberrant myeloid differentiation or the release of immature monocytes, which display higher methylation levels, and are not demethylated to the extent they are during normal differentiation.

Aberrant DNA methylation is associated with changes in gene expression of COVID-19 patient monocytes

To study the relationship between the DNA methylation changes and aberrant gene expression of monocytes derived from severe COVID-19 patients, we obtained single-cell (sc) RNA-seq data of peripheral blood mononuclear cells (PBMCs) from 10 additional severe COVID-19 patients from the same hospital and compared them with those of 10 HDs from a public dataset [41] (Additional file 2. Table S2 and Additional file 9. Figure S4A-S4B). This analysis enabled us to identify 24 cell populations based on specific markers (Fig. 4A and Additional file 9. Figure S4C-S4D), and thereby not only to determine the alterations in gene expression in monocytes, but also to inspect alterations in additional immune cell subsets. Strikingly, the monocyte fraction comprised solely CD14 + cells (CD14 mono: *CD14*) (Fig. 4B).

In the CD14 + monocyte cluster, we identified 10,440 differentially expressed genes (DEGs) between COVID-19 patients and HDs (Additional file 10. Table S6). The top DEGs (based on the fold change (FC)) included pro-inflammatory molecules *(IL1B, CCL3*), surface markers (*CD163*, *CD63*, *AREG*, *CD74*, *S100A12*, *S100A12*, *S100A8*, *S100A9*), and transcription factors (*JUN*, *MAFB*, *NF-KB*) (Fig. 4C). We observed upregulation of monocyte-derived cell markers like *S100A12*, *S100A8,* and *S100A9*. *S100A8* is already known to contribute to the cytokine storm in severe COVID-19 [41, 64]. Pro-inflammatory genes like *IL1B* of *IRF1* were downregulated, as well as *HLA* genes, in agreement with previous studies, suggesting decreased antigen presentation in severe COVID-19 patients. Finally, we observed downregulation of the NF-κB inhibitor zeta-encoding gene *NFKBIZ*, consistent with activation of this pro-inflammatory pathway [65]. Since type I IFNs are essential for antiviral immunity, and the DNA methylation analysis had indicated the potential occurrence of epigenetic alterations in IFN-stimulated genes (ISGs), we checked the expression levels of genes regulated by type I IFNs and found downregulation of several ISGs, such as *STAT1*, *BST2*, *PTPN6,* and *IRF1* (Additional file 11. Figure S5A). In addition, given that some of the observed DNA methylation changes were associated with genes involved in antigen presentation, we inspected HLA genes in our expression data and found this gene set to be significantly downregulated,

(See figure on next page.)

Fig. 3 Comparative analysis of DNA methylation in blood monocytes of severe COVID-19 and bacterial sepsis patients. **A** Violin plot representing the mean methylation state of the DMPs found in the comparison between HDs and sepsis patients with *b*-values obtained from severe COVID-19 patients. **B** Fisher's exact test showing the odds ratio ± 95% confidence interval of the overlap between DMPs found in monocytes from bacterial sepsis patients and DMPs in monocytes from COVID-19 patients. **C** Proportions of the genomic locations (in relation to genes) of DMPs in COVID-19 and sepsis; Bg., background, EPIC probes. **D** Venn diagram of the overlap of COVID-19 DMPs identified by the comparison of HDs and severe COVID-19 patients with DMPs identified by the comparison between HDs and sepsis patients, separating hypermethylated and hypomethylated DMPs. **E** Gene ontology analysis of hypermethylated and hypomethylated overlapping DMPs identified in the previous comparison. Selected significant categories ($p < 0.05$) are shown. **F** TF binding motif analysis of shared hypermethylated and hypomethylated DMPs comparing HDs and COVID-19 patients, and by HDs and sepsis patients. The panel shows the fold change (FC), TF family. Boxes with black outlines indicate TF binding motifs with FDR < 0.05. **G** Box-plot showing the DNA methylation values of individual CpGs (together with the name of the closest gene and its position relative to the transcription start site) from the hypermethylated and hypomethylated clusters from both COVID-19 and sepsis. **H** Scaled DNA methylation heatmap of regions from the whole-genome bisulfite sequencing (WGBS) data of hematopoietic stem cells (HSCs), multipotent progenitors (MPPs), common myeloid progenitors (CMPs), and granulocyte macrophage progenitors (GMPs) that overlap with the genomic location of the 1772 hypermethylated DMPs identified in the COVID-19 vs. HDs comparison. Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

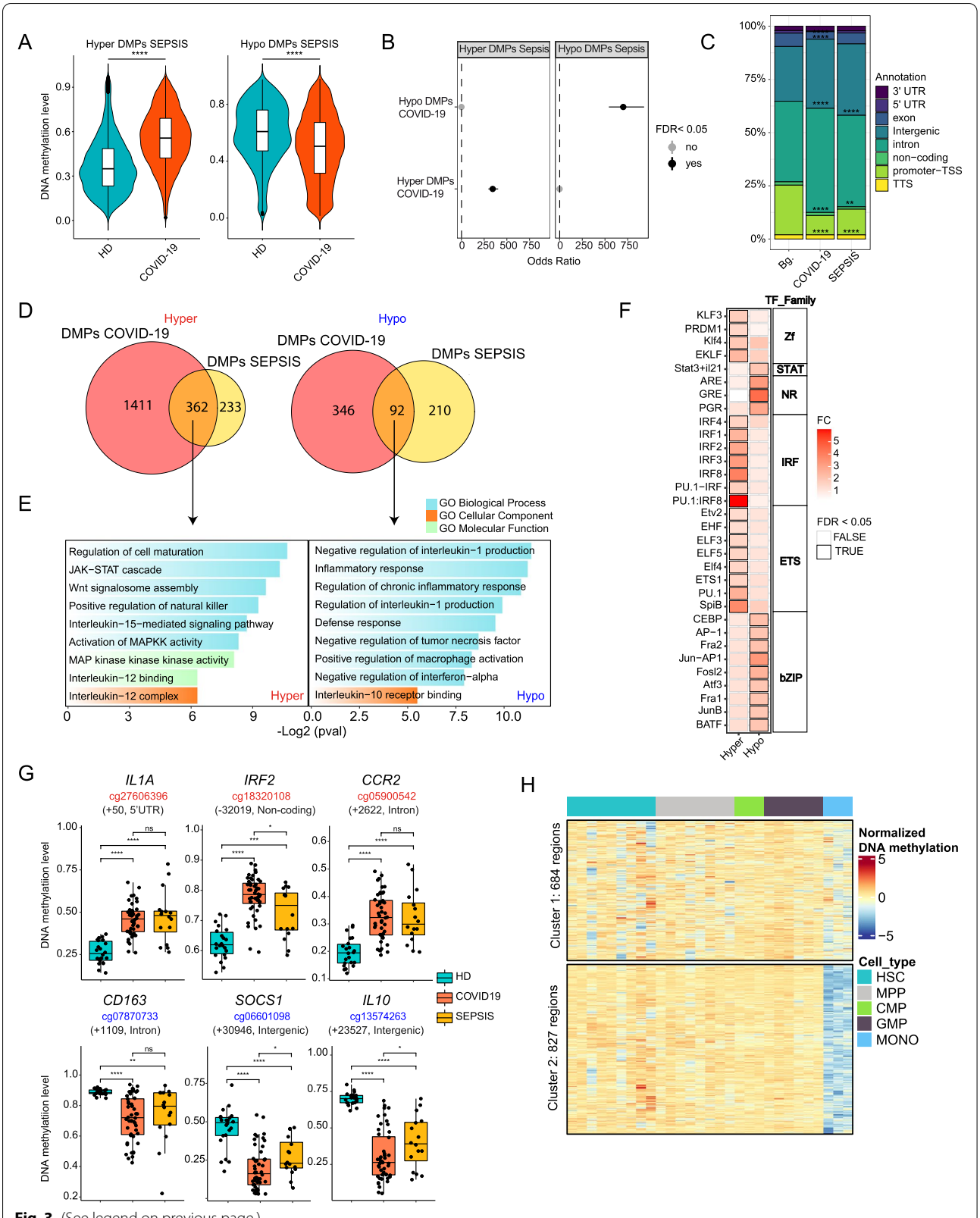


Fig. 3 (See legend on previous page.)

consistent with dysfunction in antigen processing and presentation (Additional file 11. Figure S5B).

GO analysis of both DEG sets revealed enrichment in functional terms coincident with those from DNA methylation analysis. We observed functional categories such as cytokine-mediated signaling, IL-12-mediated signaling, negative regulation of T cell activation, negative regulation of IFNγ production, and defense response to the virus in the upregulated cluster genes (Fig. 4D). Conversely, functional categories such as antigen processing and presentation by MHC-I and MHC-II and IFNγ-mediated signaling were enriched among the downregulated gene set (Fig. 4D). We then studied TFs potentially involved in the transcriptomic changes observed in COVID-19 monocytes, using Discriminant Regulon Expression Analysis (DoRothEA), and found that MAF family members, GATA3, STAT4, and IRF4, were associated with upregulated genes in severe COVID-19 (Fig. 4E). Conversely, STAT6, STAT2, IRF2, IRF3, and LYL1 were associated with downregulated genes (Fig. 4E). TF enrichment of upregulated and downregulated genes was also consistent with the results from DNA methylation analysis, in which binding motifs for several of these TFs were overrepresented among the regions neighboring the DMPs.

We determined the significance of a negative correlation between DMPs and the expression levels of their closest genes (rho = −0.31; $p = 9.8e-16$) (Fig. 4F). To study the relationship between DNA methylation and expression changes further, we performed Gene Set Enrichment Analysis (GSEA) of the genes associated with hypermethylated and hypomethylated CpG clusters. Genes associated with hypermethylated CpGs were generally downregulated (NES = 1669; FDR = 0.0005), whereas those associated with hypomethylated CpGs were upregulated (NES = −1187; FDR = 0.0596) in COVID-19 patients (Fig. 4G). GO analysis of genes with an inverse relationship between methylation and expression levels showed enrichment of functional categories like negative regulation of T cells, IFNα, and antigen presentation (Additional file 11. Figure S5C-S5D). This analysis reinforced the relationship between DNA methylation changes and expression changes related to the acquisition of a more tolerogenic phenotype in monocytes in COVID-19 patients. Some examples include *IL10*, a tolerogenic cytokine whose expression is increased in COVID-19, and *NFKBIz*, whose level of expression is decreased (Fig. 3H). We validated these results using bisulfite pyrosequencing and qRT-PCR with a new cohort of severe COVID-19 patients (Additional file 11. Figure S5E-F). The analysis also included mild COVID-19 that showed partial or total DNA methylation changes to the extent seen in severe COVID-19 cases (Additional file 11. Figure S5E-F).

Potential relationship between transcriptional and epigenetic reprogramming and altered immune cell–cell communication

Given the overrepresentation of genes associated with cytokine activity, MHC class II-mediated antigen presentation among the observed DNA methylation, and gene expression alterations in severe COVID-19, we explored the potential correlation of these changes in monocytes with their pattern of communication with other immune cell types. To systematically analyze the effect of cell–cell communication on monocytes, we used CellPhoneDB (www.cellphonedb.org), a repository of ligands, receptors, and their interactions integrated within a statistical framework that predicts enriched cellular interactions between two cell types using scRNA-seq datasets. This allowed us to infer potentially altered interactions between monocytes and other immune cell subsets in severe COVID-19. In particular, we inspected cell–cell communication alterations between CD14 + and CD4 + memory, CD4 + naïve, CD8 + memory, and CD8 + naïve T cells; B cell subsets including

(See figure on next page.)

Fig. 4 Correlation between DNA methylation and gene expression. **A** UMAP visualization showing the immune cell populations identified from Louvain clustering and cell-specific marker gene expression. **B** Dot plot representing the expression of selected marker genes identified in the cell population. The scale represents the mean gene expression level in the cell subset and the circle size represents the percentage of cells in the subset of expressing cells. **C** Heatmap representing differentially expressed genes (DEGs) with a $\log_2(FC) > 0.6$, above, and $\log_2(FC) < -0.6$, below. Genes overexpressed and downregulated in COVID-19 patients in relation to HDs are depicted in red and blue, respectively. **D** Gene ontology (GO) overrepresentation of GO Biological Process categories comprising the upregulated and downregulated DEGs. The odds ratios for each group and the $-\log_2(FC)$ are shown. Selected significant categories (FDR < 0.05) are shown. **E** Discriminant Regulon Expression Analysis (DoRothEA) of COVID-19 severe patients compared with HDs. Normalized enrichment score (NES) and $\log_2(FC)$ of transcription factor expression are depicted. **F** Correlation of average DNA methylation levels of DMPs with average gene expression of DEGs in the HDs *vs.* COVID-19 severe patients. $\text{Log}_2(FC)$ of expression is plotted on the *y*-axis, higher numbers representing a higher level of expression in COVID-19 and lower numbers a higher level of expression in HDs. DNA methylation is depicted on the *x*-axis as Δβ, lower numbers representing a lower level of methylation in COVID-19 monocytes, and higher numbers a lower level of methylation in HDs. Points are colored according to their genomic context. **G** Gene set enrichment analysis (GSEA) of HD *vs.* COVID-19, using hypomethylated-associated genes and hypermethylated-associated genes as genesets. The running enrichment score is represented, and the normalized enrichment score (NES) is shown above (FDR < 0.01). **H** Representation of individual DNA methylation values of DMPs from the hypermethylated and hypomethylated clusters (*beta* values), the position in respect to the transcription start site, and the relative expression of the closely related DEGs. Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$, **** $p < 0.00001$

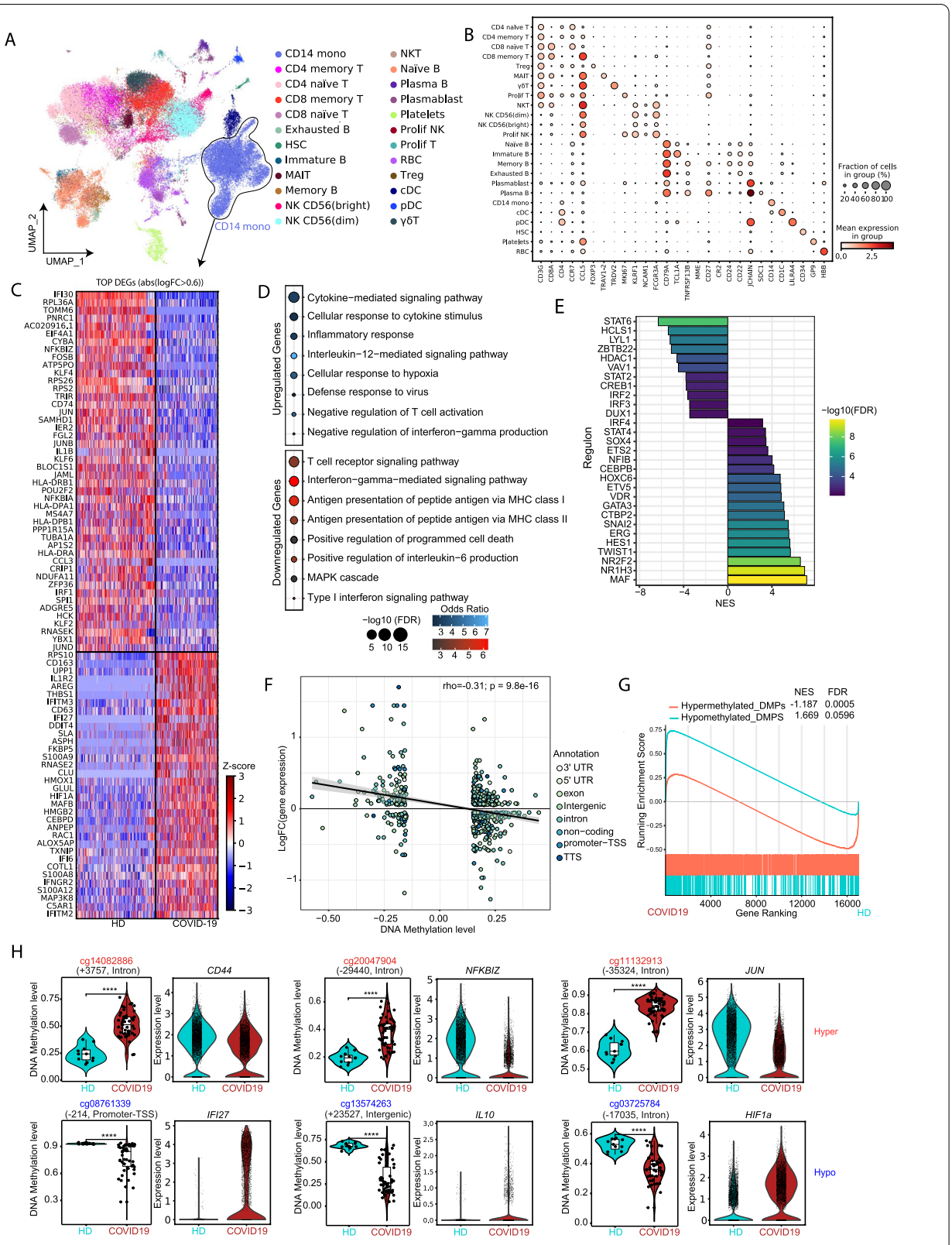


Fig. 4 (See legend on previous page.)

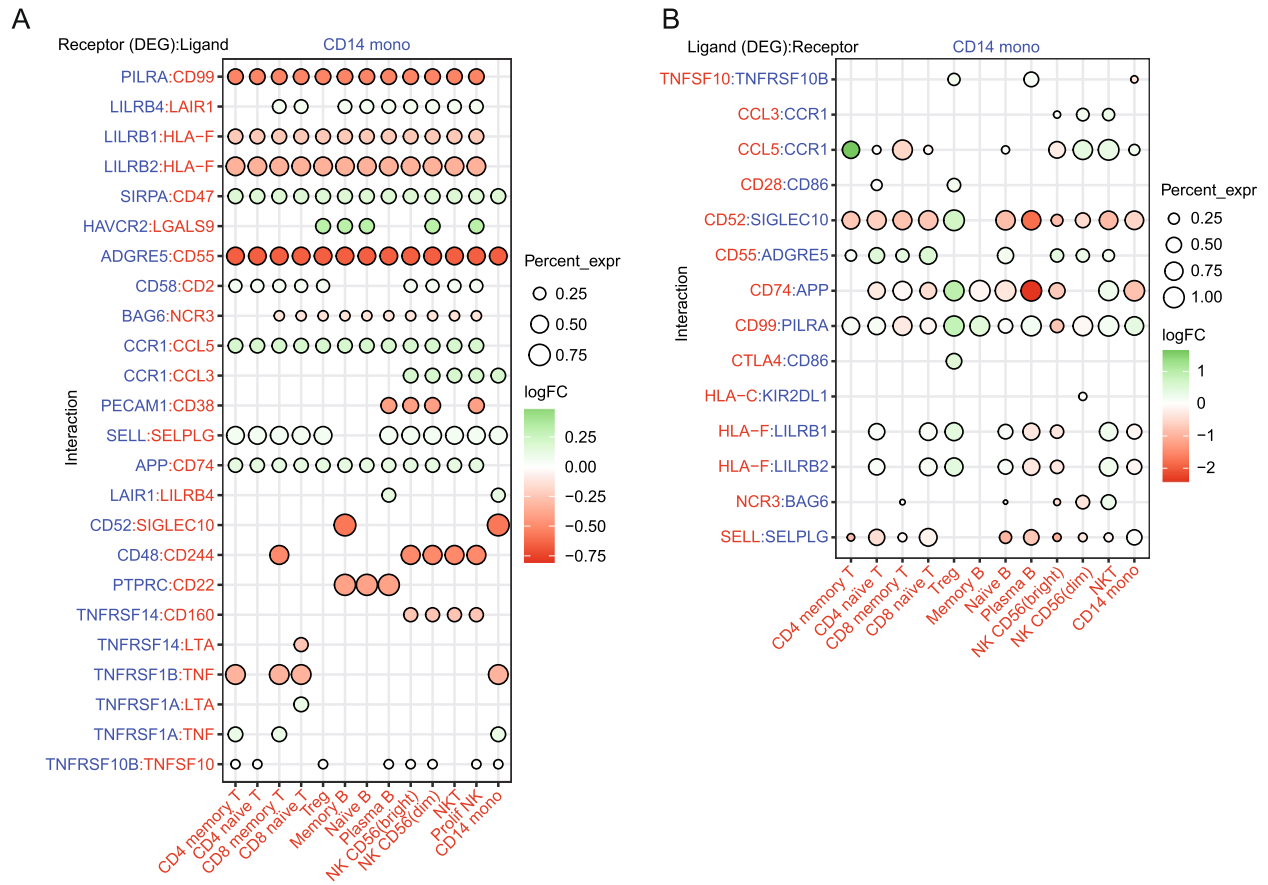


Fig. 5 Cell–cell communication analysis. Dot plot of selected receptor/ligand pair (**A**) and ligand/receptor (**B**) interactions between CD14 + monocytes and other cell components in the COVID-19 patient group. Gene expression is indicated as $\log_2$(FC) for differentially expressed genes (FDR < 0.05), which, in both cases (**A** and **B**), are the molecules presented on the left. The percentage expression of the differentially expressed genes in each cell type is indicated by the circle size. Molecules shown in blue are those expressed in CD14 + monocytes. Molecules expressed in the immune cell partner are shown in red

memory, naïve, and plasma B cells; natural killer cells (NK $CD56^{dim}$: NK $CD56^{bright}$) (Fig. 5A,B). Our analysis revealed 4483 ligand/receptor pairs, in which the expression levels of ligands and receptors of CD14 + and/or interacting partners in the aforementioned cell types were significantly different between severe COVID-19 patients and HDs, suggesting changes in the interaction of the corresponding immune cells (Additional file 12. Table S7). The aberrant levels of the proteins encoded by these genes in monocytes were validated by flow cytometry (Figure S5G), supporting a potential impact on cell–cell communication.

Figure 5A illustrates the significant ligand-receptor interactions that may be affected when the expression of receptor in monocytes is altered, revealing their potential impact on other cell types. In general, there was a high frequency of interactions involving different types of NK cells, consistent with the terms observed in the GO analysis performed with DMPs (Fig. 1E). *PILRA*, *LILRB1*, *LILRB2*, and *PECAM1* (CD31), the products of which are involved in the inhibition of immune response, were downregulated in monocytes. Their corresponding ligand-encoding genes, *CD99*, *HLA-F*, and *CD38*, were expressed in all the analyzed cell types, except for CD38, which is only expressed in NK and plasma B cells. Additionally, the gene encoding for receptor LAIR1, which inhibits IL-2 expression, was upregulated in monocytes [66], which might influence the interaction with cells expressing its corresponding ligand, i.e., plasma B cells and monocytes. Our analysis also revealed changes in the expression of TNF receptor genes (*TNFRSF14*, *TNFRSF1B*, *TNFRSF1A*) in monocytes, which could affect the interaction with T cells through the ligands encoded by *TNF* and *LTA*. This is compatible with the possibility that TNF-associated DNA methylation alterations in monocytes could arise from altered interactions with T cells through these ligand-receptor pairs. We also noted downregulation of the receptor *TNFRSF14*, which

interacts with CD160 in NK cells. Some studies have argued that CD160 is essential for NK-mediated IFNγ production [67], a conclusion that is consistent with the results obtained in our gene ontology analysis of the DNA methylation data. *ADGRE5* (CD97) was downregulated in monocytes. This receptor interacts with CD55, which is expressed in all the analyzed cell types. This interaction is involved in leukocyte migration [68]. The potential alteration of this interaction could be linked to the observed hypomethylation of CpGs close to genes related to leukocyte migration (Fig. 1E, top).

We also examined DEGs corresponding to ligands expressed in all immune cell types, whose corresponding receptors are expressed in monocytes, to identify potential cell–cell communication alterations that might affect monocytes (Fig. 5B). In general, we detected upregulation of ligands in regulatory T cells (Treg) and downregulation of ligands in plasma B cells. We also observed increased levels of *CCL5* and *CCL3*, expressed in NK cells, that interact with the CD191 receptor (*CCR1*), and whose inhibition potentially suppresses immune hyperactivation in critical COVID-19 patients [69]. In the context of antigen presentation, there was upregulation of HLA-F from Treg and NKT, which interacts with *LILRB1* in monocytes. Recent studies have associated *LILRB1* with the development of tolerance [70]. Our analysis also revealed low levels of CD99, expressed in CD4 + memory and naïve T cells, Treg and memory B cells, and the receptor PILRA, which is expressed in monocytes. The opposite occurs with CD8 memory and naïve T cells and NK CD56(bright), which enhances T cell migration [71]. There was a similar trend between CD74 and the receptor APP expressed in monocytes, which is involved in antigen processing and presentation. This could be related to the impaired antigen presentation previously highlighted in our data.

In brief, the potential alteration of cell–cell communication events, through increased or decreased levels of ligands and receptors involving inflammatory cytokines, antigen presentation-related factors, and cell activation regulators, in severe COVID-19 patients could affect downstream cell-signaling pathways and TFs and perhaps influence DNA methylation profiles in monocytes, thereby perpetuating aberrant immune responses.

Discussion

Our results reveal that peripheral blood monocytes from severe COVID-19 patients display aberrant DNA methylomes and transcriptomes associated with functions related to IFN type I signaling and antigen presentation, among others. The changes are significantly associated with organ damage and with DNA methylation changes occurring in bacterial sepsis. Finally, our analysis suggests that pro-inflammatory cytokines, the release of immature or aberrant monocytes, and specific dysregulated immune cell–cell communication events may be responsible for some epigenetic changes.

To date, there have been very few DNA methylation studies addressing the involvement of COVID-19 DNA methylation in regulating the angiotensin-converting enzyme 2 (ACE2) type I membrane receptor gene [72], which is present in arterial, lung type II alveolar cells, where it acts as a SARS-CoV-2 receptor. There is a suggestion that the host epigenome may represent a risk factor for COVID-19 infection. Very few studies have reported alterations in DNA methylation in relation to immune responses [73–75]. Our study aimed to explore the involvement of DNA methylation in relation to a severe COVID-19 outcome in the myeloid compartment, which is directly related to systemic inflammation. We specifically studied monocytes because it is the cell type that undergoes the most dramatic transcriptomic reprogramming during COVID-19 infection [13, 21, 23, 76]. In this regard, our study provides the first instance of DNA methylome profiling in a specific immune cell type in COVID-19 patients.

Our data revealed that most DNA methylation changes in monocytes derived from severe COVID-19 patients occurred in genomic sites enriched in PU.1 binding motifs, consistent with earlier studies showing its role as a pioneer TF directly recruiting TET2 and DNMT3b [77]. In our case, most DNA methylation changes occurred in genes related to cytokines, MHC class II proteins, and IFN signaling. Similar results about the defective function of MHC-II molecules and activation of apoptosis pathways were obtained in single-cell atlas studies of PBMCs from severe COVID-19 patients [6, 21, 78, 79] and in sepsis [80, 81].

We found that DNA methylation changes in severe COVID-19 patients share some features with sepsis, especially those associated with the expression of tolerogenic cytokines like IL-10 [82]. The acute phase of these infections suggests a dysregulated inflammatory host response, resulting in an imbalance between pro-inflammatory and anti-inflammatory mediators [14]. Some studies have suggested that viral components induce STAT1 dysfunction and compensatory hyperactivation of STAT3 in SARS-CoV-2-infected cells [83]. We noted the involvement of kinases like JNK, and earlier studies had shown that COVID-19 infection activates the *JNK* and *ERK* pathways that end in the AP-1-dependent gene expression of pro-inflammatory cytokines [84]. One of the most strongly affected TFs is STAT2, together with STAT6, which could be linked to the aberrant IFN signaling in monocytes in COVID-19 [83]. The presence of STAT2 downregulation also

suggests a deficiency in the ability to cross-present to CD8 + T cells [85].

We also identified GRE binding sites in association with DNA methylation changes. Generally, the glucocorticoid receptor (GR) is activated when patients are treated with GC. However, we also noted significant GRE enrichment in patients who were not treated with GC, suggesting that endogenous production of GC in COVID-19 patients could regulate GR and affect DNA methylation at its genomic binding sites. GC is also produced endogenously in sepsis patients in whom cytokines like IL-1β, TNFα, and IL-6 induce its production from the adrenal cortex using cholesterol as a substrate to reduce inflammatory responses [86, 87]. These cytokines were hypomethylated and overexpressed in our dataset, consistent with the results of other studies that have reported increased levels in the serum of COVID-19 patients [88, 89]. GRE binding sites are enriched in the DMPs common to COVID-19 and sepsis. GR is a nuclear receptor expressed in most cell types that can trigger the expression of anti-inflammatory genes through direct DNA binding. Furthermore, GRE represses the action of other inflammation-related TFs, including members of the NF-KB and AP-1 families [90, 91], which are also known to be downregulated in our cohort. Taken together, our results suggest the existence of a relationship between extracellular factors associated with the cytokine storm occurring in severe COVID-19 and DNA methylation changes. Several studies have shown an increase in the levels of inflammatory cytokines in severe COVID-19, which may contribute to the severity of the disease [92].

However, it is also possible that the DNA methylation changes are partly due to the release of immature or altered monocytes from myelopoiesis, as reported for severe COVID-19 [13, 20, 93, 94] and sepsis [63]. Release of immature myeloid cells from the bone marrow in severe COVID-19 is reminiscent of emergency myelopoiesis [95]. This is a well-known phenomenon, characterized by the mobilization of immature myeloid cells to restore functional immune cells, and by its contribution to the dysfunction of innate immunity [96]. In fact, a proportion of the hypermethylated CpGs in monocytes from severe COVID-10 patients overlap with regions that become demethylated during myeloid differentiation. This suggests that part of the hypermethylated CpG sites in isolated peripheral blood CD14 + might be associated with aberrantly differentiated monocytes released into the bloodstream in severe COVID-19 patients. However, the small numbers of CD34 + cells in the PBMC fraction of COVID-19 patients and the lack of CD14 + cells in this subset suggest no interference with our results for CD14 + CD15- cells, isolated with our method.

The relationship between DNA methylation and gene expression is complex. DNA methylation patterns are cell-type-specific and are established during dynamic differentiation events by site-specific remodeling at regulatory regions [97]. In general, methylation of CpGs located in gene promoters, first exons, and introns is negatively correlated with gene expression [98]. The analysis of our data shows that there is an inverse correlation between the CpG methylation changes and the expression levels of the closest genes. The comparison of the inferred TFs associated with DNA methylation changes and gene expression changes shows common factors like IRF2 and IRF3, which regulate downregulated genes and hypermethylated CpGs. In this context, it is possible that reduced levels of IFN regulatory factor IRF3 or defective IRF7 function reduces the level of IFNα/β gene expression, increasing the sensitivity to viral infection [12, 99].

Finally, analysis of cell–cell communication has revealed potential relationships between DNA methylation changes and altered communication of monocytes and other immune cells (e.g., T, plasma B and NK cells). Our data suggest the potential reduction of interactions between monocytes and NK cells through CD160, which mediates the antibody-dependent cell-mediated cytotoxicity that it is essential for IFNγ production [67]. The potentially greater interaction between monocytes and Treg through multiple ligand and receptor pairs is an interesting finding, since Tregs are immunosuppressive cells responsible for maintaining immune homeostasis [100]. In any case, the use of CellPhone DB is useful for inferring cell–cell communications events; however, additional validation experiments would be necessary to validate interactions and activation of downstream signaling pathways.

In our study, we could not determine whether the observed DNA methylation alterations in COVID-19 were the cause or the consequence of the changes in gene expression. The analysis of mild COVID-19 cases, in which the DNA methylation and expression level of a few genes showed differences in their similarities with severe COVID-19 cases, suggests that there are cases where expression changes might anticipate DNA methylation changes. In any case, it is reasonable to propose that some DNA methylation changes help perpetuate dysregulated immune responses.

Some limitations of our study include the size of the cohort, and the unequal numbers of individuals administered particular drugs in the different patient groups, which could have affected the COVID-19 data. However, despite these limitations, we found no significant differences among severe COVID-19 patients with respect to the time they were admitted to the ICU or began to receive treatment. This suggests that DNA

methylation is quite a general occurrence in the context of COVID-19. Another limitation concerns the cell population analyzed, since the method for monocyte isolation comprises two populations, CM and IM, one of which (CM) is expanded in the patient group. However, the analysis including the monocyte subsets as a covariate indicates that there are no major differences. Finally, in the comparison with DNA methylation of progenitor cells, it is important to note that the DMPs were overlapped with genomic regions, and not single-base data, and further analyses would be required.

Future studies would benefit from having access to a wider cohort in which it is possible to identify significant links between alterations and drug treatments. Incorporating mild and asymptomatic cases would improve our ability to dissect drug- and severity-related specificity in relation to DNA methylation changes. As is the case for other medical conditions, the analysis 1of DNA methylation changes would be very likely to help predict disease severity, progression, and recovery.

Conclusions

Our study provides unique insights into the epigenetic alterations of monocytes in severe COVID-19. We have shown that peripheral blood monocytes from severe COVID-19 patients undergo changes in their DNA methylomes, in parallel with changes in expression, and that these significantly overlap with those found in patients with sepsis. We have also shown DNA methylation changes are associated with organ dysfunction. Finally, our results suggest a relationship between DNA methylation changes in COVID-19 patients and changes that occur during myeloid differentiation and others that can be induced by pro-inflammatory cytokines. CellPhoneDB analysis also suggests that alterations in immune cell crosstalk can contribute to transcriptional reprogramming in monocytes, which involves dysregulation of interferon-related genes and genes associated with antigen presentation and chemotaxis.

Abbreviations

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; COVID-19: Coronavirus disease 2019; IL: Interleukin; GM-CSF: Granulocyte-macrophage colony-stimulating factor; TNF-a: Tumor necrosis factor-alpha; IFN: Interferon; IRF: Interferon regulatory factor; NK: Natural killer; DC: Dendritic cell; MHC-II: Class II major histocompatibility complex; 5mC: 5-Methylcytosine; TF: Transcription factor; DNMT: DNA methyltransferase; TET: Ten–eleven translocation; scRNA-seq: Single-cell RNA sequencing; ICU: Intensive care unit; PBMC: Peripheral blood mononuclear cell; HD: Healthy donor; FBS: Fetal bovine serum; BS: Bisulfite; FDR: False discovery rate; DVPs: Differentially variable positions; PCA: Principal component analysis; PC: Principal component; HSC: Hematopoietic stem cell; MPP: Multipotent progenitor; CMP: Common myeloid progenitor; GMP: Granulocyte macrophage progenitor; WGBS: Whole-genome bisulfite sequencing; UMAP: Uniform Manifold Approximation and Projection; DoRothEA: Discriminant Regulon Expression Analysis; NES: Normalized enrichment score; RT-qPCR: Real-time quantitative Polymerase Chain Reaction; CM: Classical monocytes; IM: Intermediate monocytes; NCM: Non-classical monocytes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-022-01137-4.

Additional file 1: Table S1. Clinical features of severe COVID-19 patients and healthy donors (HD) included in the DNA methylation analysis.

Additional file 2: Table S2. Clinical features of severe COVID-19 patients and healthy donors (HD) included in scRNAseq analysis.

Additional file 3: Table S3. Primer sequences.

Additional file 4: Figure S1. Flow cytometry profiles of a representative sample for each group; HD (**A**) and COVID-19 (**B**) indicating the sorting strategy and gates used in the study. **C** Boxplot representing the mean percentage of CD14+ and CD15+ cells in HD and patients. **D** Flow cytometry profile of the CD14+CD15- purified monocytes using the same gates used in the study. **E** HD and COVID-19 indicating the gates used for monocyte subtype analysis (classical monocytes, CM; intermediate monocytes, IM; and non-classical monocytes, NCM) (**F**) Boxplot representing the mean percentage of CM, IM and NCM in HD and COVID-19 patients.

Additional file 5: Table S4. List of hypermethylated and hypomethylated genes in COVID-19 monocytes (FDR < 0.05; $\Delta\beta \geq 0.2$).

Additional file 6: Figure S2. DNA methylation analysis in blood monocytes of severe COVID-19 patients. **A** Venn diagram of the overlap of hyper- and hypomethylated DMPs identified with a subcluster of samples including (CM/IM_YES) or not (CM/IM_NO) the CM and IM percentage as a covariable in the comparison between HD and severe COVID-19 patients (**B**) Heatmap representation of beta values of first 10 Principal component (PC) analysis that correlations PCs with different clinical variables. Numerical variables were correlated to PCs using Pearson correlation, whereas categorical variables were entered in a linear model together with the PCs. **C** Principal component analysis (PCA) of the DMPs. The HDs are illustrated in grey, and the severe COVID-19 patients are illustrated as blue and red in function of the clinical parameter or treatment with dexamethasone. **D** Enrichment analysis of different chromatin states for CpGs sites corresponding the Hyper- and Hypomethylated clusters. The FDR is represented with the size of the bubble, as shown. The relative enrichment is represented as Odds Ratio. TssA, Active TSS; TssBiv, Bivalent/Poised TSS; BivFlnk, Flanking Bivalent TSS/Enh; EnhBiv, Bivalent Enhancer; ReprPC, Repressed PolyComb; ReprPCWk, Weak Repressed PolyComb; Quies, Quiescent/Low; TssAFlnk, Flanking Active TSS; TxFlnk, Transcr. at gene 5′ and 3′; Tx, Strong transcription; TxWk, Weak transcription; EnhG, Genic enhancers; Enh, Enhancers; ZNF/Rpts, ZNF genes & repeats; Het, Heterochromatin. **E** TF binding motif analysis of hypomethylated DMPs comparing patients no treated with dexamethasone vs. HDs (DEX_no) and patients treated with dexamethasone vs. HDs (DEX_yes). The panel shows the fold change (FC), TF family. Black outlined boxes indicate TF binding motifs with FDR values < 0.05. **F** Box plot of individual DNA methylation values of CpG from the hypermethylated and hypomethylated clusters with the name of the closest gene and the position in respect to the transcription start site calculated using pyrosequencing in the validation cohort that include HD and patients with mild and severe infection of COVID-19. **G** Volcano plot showing the p value vs the variance ratio for HD and COVID-19 associated differentially variable CpG positions (DVPs). DVPs were identified using the algorithm iEVORA. Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$, **** $p < 0.00001$.

Additional file 7: Table S5. List of hypermethylated and hypomethylated CpGs related to SOFA (rho < 0.4/ rho > 0.4 and $p < 0.01$).

Additional file 8: Figure S3. DNA methylation comparative analysis between blood monocytes of severe COVID-19 patients and bacterial sepsis patients (**A**) Proportions of the genomic locations (in relation to genes) of hyper- and hypomethylated DMPs in COVID-19 and sepsis; Bg., background, EPIC probes. **B** Venn diagram of the overlap of COVID-19

DMPs identified by the comparison between HD and severe COVID-19 patients with DMPs identified by the comparison between HD and septic patients. Gene ontology (GO) analysis of the shared (**C**) and not shared (**D**) DMPs from the previous representation. **E** Violin plot of the mean methylation status of the identified DMPs with b-values obtained from monocytes derived from healthy donor PBMCs exposed in vitro for 4 days to interferon-alpha (IFNα) (100ng/mL), interferon gamma (IFNγ) (100ng/mL), tumor necrosis factor-alpha (TNFα) (10ng/mL) and untreated (Control) (n=3), [26]. Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$, **** $p < 0.00001$.

Additional file 9: Figure S4. A UMAP visualization showing the two groups identified from Louvain clustering. **B** Barplot representation of the proportion of the two study groups in the different cell type found. **C** Barplot representation of proportion of each cell type in each study group. **D** Barplot representation of proportion of each cell type in each sample used for the scRNA-seq analysis.

Additional file 10: Table S6. List of Differentially Expressed Genes (DEGs) in CD14+ cells between severe COVID-19 patients and HD (FDR < 0.05).

Additional file 11: Figure S5. Analysis of DEG in monocytes derived from severe COVID-19 patients. **A** Heatmap representation of expression levels of IFN genes (ligands and receptor) in HD and COVID-19 CD14+ monocytes. Gene ontology (GO) analysis of the upregulated. **B** Dot plot of major histocompatibility complex (MHC) genes. LogFc and percentage of expression is representeate. **C** and downregulated (**D**) DEG that present a negative correlation with their close CpG. **E** Box plot of individual DNA methylation values of CpG from the hypermethylated and hypomethylated clusters with the name of the closest gene and the position in respect to the transcription start site. Calculated using pyrosequencing in the validation cohort that include HD and patients with mild and severe infection of COVID-19. **F** Box plot of relative expression of individual genes performed by real-time quantitative polymerase chain reaction (RT-qPCR) in the validation cohort that include HD and patients with mild and severe infection of COVID-19. **G** Box plot of mean florescence intensity (MFI) of cell surface markers in CD14+ cells in the cohort used for single cell analysis. Statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$, **** $p < 0.00001$.

Additional file 12: Table S7. List of ligand-receptor pairs with different expression in COVID-19.

Acknowledgements
We thank the CERCA Programme/Generalitat de Catalunya and the Josep Carreras Foundation for institutional support. We are very grateful to the Cytometry Laboratory and the Scientific and Technological Centers, University of Barcelona (CCiT-UB), for their help with the gating strategy and cell sorting.

Authors' contributions
G.G.-T. and E.B. conceived and designed the study; G.G.-T., A.G.F.-B., and L.C. prepared and purified the samples; G.G.-T., O.M.-P., C.C.-F., and R.M.-F. performed bioinformatic analyses; A.B. analyzed the single-cell datasets; A.R.-S., M.M.G., R.F., and J.C.R.-R. provided the patient samples and analyzed the clinical data; G.G.-T., J.R.-U., R.V.-T., and E.B. analyzed and interpreted the data; G.G.-T. and E.B. wrote the manuscript; all authors read and approved the final manuscript.

Funding
This study has been funded by R+D+i project of the Spanish Ministry of Science and Innovation (grant number PID2020-117212RB-I00/ MICIN/ AEI/10.13039/501100011033). We also thank the Chan Zuckerberg Initiative (grant 2020–216799) and Wellcome Sanger core funding (WT206194). This publication has also been supported by the Unstoppable campaign of the Josep Carreras Leukaemia Foundation. Additional funding comes from Project PI18/00346 (Instituto de Salud Carlos III and co-funded by European Union (ERDF/ESF). A.B. received additional support from a Gates Cambridge Scholarship.

Availability of data and materials
DNA methylation data for this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible through GEO SuperSeries accession number GSE188573 [30]. Single-cell data generated in this study are available as h5ad files at the European Genome-Phenome Archive (EGA) with Dataset ID EGAD00001007982 [101].

Declarations

Ethics approval and consent to participate
This study was approved by the Clinical Research Ethics Committees of Hospital Universitari Germans Trias i Pujol (PI-20–129) and Vall d'Hebron University Hospital (PR(AG)282/2020), which adhered to the principles set out in the WMA Declaration of Helsinki. All samples were managed in compliance with participants' written informed consent to participate in the study.

Consent for publication
Written informed consent for publication was provided by the participants.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain. [2]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1RQ, UK. [3]Department of Medical Genetics, University of Cambridge, Cambridge CB2 0QQ, UK. [4]Intensive Care Department, Vall d'Hebron University Hospital, Shock, Organ Dysfunction and Resuscitation (SODIR) Research Group, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, 08035 Barcelona, Spain. [5]Immunology Division, Vall d'Hebron University Hospital and Diagnostic Immunology Research Group, Vall d'Hebron Research Institute (VHIR), 08035 Barcelona, Spain. [6]Epigenetics in Inflammatory and Metabolic Diseases Laboratory, Health Science Center (HSC), East China Normal University (ECNU), Shanghai 200241, China.

Received: 29 April 2022 Accepted: 7 November 2022
Published online: 29 November 2022

References

1. Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents. 2020;55:105924.
2. Schultze JL, Aschenbrenner AC. COVID-19 and the human innate immune system. Cell. 2021;184:1671–92.
3. Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. Cell. 2021;184:861–80.
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395:497–506.
5. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol. 2020;20:363–74.
6. Zhang J-Y, Wang X-M, Xing X, Xu Z, Zhang C, Song J-W, et al. Single-cell landscape of immunological responses in patients with COVID-19. Nat Immunol. 2020;21:1107–18.
7. Zhang C, Shi L, Wang F-S. Liver injury in COVID-19: management and challenges. Lancet Gastroenterol Hepatol. 2020;5:428–30.
8. Zhou Y, Fu B, Zheng X, Wang D, Zhao C, Qi Y, et al. Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients. Natl Sci Rev. 2020;7:998–1002.
9. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med. 2020;58:1021–8.
10. Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, et al. Immunophenotyping of COVID-19 and influenza highlights the role of type

I interferons in development of severe COVID-19. Sci Immunol. 2020;5:eabd1554.
11. Gustine JN, Jones D. Immunopathology of hyperinflammation in COVID-19. Am J Pathol. 2021;191:4–17.
12. Sato M, Suemori H, Hata N, Asagiri M, Ogasawara K, Nakao K, et al. Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN-alpha/beta gene induction. Immunity. 2000;13:539–48.
13. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell. 2020;182:1419-1440.e23.
14. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Smith N, et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science. 2020;369:718–24.
15. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell. 2020;181:1036-1045.e9.
16. Trouillet-Assant S, Viel S, Gaymard A, Pons S, Richard J-C, Perret M, et al. Type I IFN immunoprofiling in COVID-19 patients. J Allergy Clin Immunol. 2020;146:206-208.e2.
17. Bastard P, Rosen LB, Zhang Q, Michailidis E, Hoffmann H-H, Zhang Y, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science. 2020;370:eabd4585.
18. Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science. 2020;370:eabd4570.
19. Arunachalam PS, Wimmers F, Mok CKP, Perera RAPM, Scott M, Hagan T, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science. 2020;369:1210–20.
20. Silvin A, Chapuis N, Dunsmore G, Goubet A-G, Dubuisson A, Derosa L, et al. Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. Cell. 2020;182:1401-1418.e18.
21. Guo C, Li B, Ma H, Wang X, Cai P, Yu Q, et al. Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm. Nat Commun. 2020;11:3924.
22. Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat Med. 2020;26:1070–6.
23. Saichi M, Ladjemi MZ, Korniotis S, Rousseau C, Ait Hamou Z, Massenet-Regad L, et al. Single-cell RNA sequencing of blood antigen-presenting cells in severe COVID-19 reveals multi-process defects in antiviral immunity. Nat Cell Biol. 2021;23:538–51.
24. Bader SM, Cooney JP, Pellegrini M, Doerflinger M. Programmed cell death: the pathways to severe COVID-19? Biochem J. 2022;479:609–28.
25. Álvarez-Errico D, Vento-Tormo R, Sieweke M, Ballestar E. Epigenetic control of myeloid cell differentiation, identity and function. Nat Rev Immunol. 2015;15:7–17.
26. Rodríguez-Ubreva J, de la Calle-Fabregat C, Li T, Ciudad L, Ballestar ML, Català-Moll F, et al. Inflammatory cytokines shape a changing DNA methylome in monocytes mirroring disease activity in rheumatoid arthritis. Ann Rheum Dis. 2019;78:1505–16.
27. Lorente-Sorolla C, Garcia-Gomez A, Català-Moll F, Toledano V, Ciudad L, Avendaño-Ortiz J, et al. Inflammatory cytokines and organ dysfunction associate with the aberrant DNA methylome of monocytes in sepsis. Genome Med. 2019;11:66.
28. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: in the right place at the right time. Science. 2018;361:1336–40.
29. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011;25:1010–22.
30. Godoy-Tena G, Barmada A, Morante-Palacios O, de la Calle-Fabregat C, Ciudad L, Ruiz-Sanmartín A, Martínez-Gallo M, Frank-Bertoncelj M, Ferrer R, Ruiz-Rodriguez JC, Rodríguez-Ubreva J, Vento-Tormo R BE. Epigenetic and transcriptomic remodeling in monocytes of severe COVID-19 patients mirrors systemic changes in cytokines and immune cell crosstalk. Datasets. Gene Expression Omnibus. 2022. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE188573.
31. Morante-Palacios O, Ballestar E. shinyÉPICo: a graphical pipeline to analyze Illumina DNA methylation arrays. Bioinformatics. 2021;37:257–9
32. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
33. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010;26:2363–7.
34. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat Commun. 2016;7:10478.
35. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.
36. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28:1045–8.
37. Farlik M, Halbritter F, Müller F, Choudry FA, Ebert P, Klughammer J, et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. Cell Stem Cell. 2016;19:808–22.
38. Rodríguez-Ubreva J, Arutyunyan A, Bonder MJ, Del Pino-Molina L, Clark SJ, de la Calle-Fabregat C, et al. Single-cell Atlas of common variable immunodeficiency shows germinal center-associated epigenetic dysregulation in B-cell responses. Nat Commun. 2022;13:1779.
39. Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. Nat Methods. 2020;17:615–20.
40. Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. Cell Syst. 2019;8:281-291.e9.
41. Stephenson E, Reynolds G, Botting RA, Calero-Nieto FJ, Morgan MD, Tuong ZK, et al. Single-cell multi-omics analysis of the immune response in COVID-19. Nat Med. 2021;27:904–16.
42. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15:1053–8.
43. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20.
44. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19:15.
45. Yao C, Bora SA, Chen P, Goodridge HS, Gharib SA. Sample processing and single cell RNA-sequencing of peripheral blood immune cells from COVID-19 patients. STAR Protoc. 2021;2:100582.
46. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
47. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019;29:1363–75.
48. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. Cell PhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc Springer, US. 2020;15:1484–506.
49. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007;23:1289–91.
50. Agashe C, Chiang D, Grishin A, Masilamani M, Jones SM, Wood RA, et al. Impact of granulocyte contamination on PBMC integrity of shipped blood samples: Implications for multi-center studies monitoring regulatory T cells. J Immunol Methods. 2017;449:23–7.
51. Kapellos TS, Bonaguro L, Gemünd I, Reusch N, Saglam A, Hinkley ER, et al. Human monocyte subsets and phenotypes in major chronic inflammatory diseases. Front Immunol. 2019;10:2035.
52. Kuhns DB, Priel DAL, Chu J, Zarember KA. Isolation and functional analysis of human neutrophils. Curr Protoc Immunol. 2015;111:7.23.1–7.23.16.
53. Gatti A, Radrizzani D, Viganò P, Mazzone A, Brando B. Decrease of non-classical and intermediate monocyte subsets in severe acute SARS-CoV-2 infection. Cytometry A. 2020;97:887–90.
54. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13:484–92.

55. Ribas VJ, Lopez JC, Ruiz-Rodriguez JC, Ruiz-Sanmartin A, Rello J, Vellido A, On the use of decision trees for ICU outcome prediction in sepsis patients treated with statins. IEEE Symp Comput Intell Data Min. IEEE. 2011;2011:37–43.
56. López-Collazo E, Avendaño-Ortiz J, Martín-Quirós A, Aguirre LA. Immune response and COVID-19: a mirror image of sepsis. Int J Biol Sci. 2020;16:2479–89.
57. Pietras EM, Saha SK, Cheng G. The interferon response to bacterial and viral infections. J Endotoxin Res. 2006;12:246–50.
58. Liu Q, Ou Q, Chen H, Gao Y, Liu Y, Xu Y, et al. Differential expression and predictive value of monocyte scavenger receptor CD163 in populations with different tuberculosis infection statuses. BMC Infect Dis. 2019;19:1006.
59. Townsend L, Dyer AH, Naughton A, Imangaliyev S, Dunne J, Kiersey R, et al. Severe COVID-19 is characterised by inflammation and immature myeloid cells early in disease progression. Heliyon. 2022;8:e09230.
60. Gibellini L, De Biasi S, Paolini A, Borella R, Boraldi F, Mattioli M, et al. Altered bioenergetics and mitochondrial dysfunction of monocytes in patients with COVID-19 pneumonia. EMBO Mol Med. 2020;12:e13001.
61. Brauns E, Azouz A, Grimaldi D, Xiao H, Thomas S, Nguyen M, et al. Functional reprogramming of monocytes in patients with acute and convalescent severe COVID-19. JCI insight. 2022;7.
62. Kvedaraite E, Hertwig L, Sinha I, Ponzetta A, Hed Myrberg I, Lourda M, et al. Major alterations in the mononuclear phagocyte landscape associated with COVID-19 severity. Proc Natl Acad Sci U S A. 2021;118:e2018587118.
63. Reyes M, Filbin MR, Bhattacharyya RP, Sonny A, Mehta A, Billman K, et al. Plasma from patients with bacterial sepsis or severe COVID-19 induces suppressive myeloid cell production from hematopoietic progenitors in vitro. Sci Transl Med. 2021;13:eabe9599.
64. Ren X, Wen W, Fan X, Hou W, Su B, Cai P, et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. Cell. 2021;184:1895-1913.e19.
65. Hariharan A, Hakeem AR, Radhakrishnan S, Reddy MS, Rela M. The role and therapeutic potential of NF-kappa-B pathway in severe COVID-19 patients. Inflammopharmacology. 2021;29:91–100.
66. Spolski R, Li P, Leonard WJ. Biology and regulation of IL-2: from molecular mechanisms to human therapy. Nat Rev Immunol. 2018;18:648–59.
67. Tu TC, Brown NK, Kim T-J, Wroblewska J, Yang X, Guo X, et al. CD160 is essential for NK-mediated IFN-γ production. J Exp Med. 2015;212:415–29.
68. Safaee M, Clark AJ, Ivan ME, Oh MC, Bloch O, Sun MZ, et al. CD97 is a multifunctional leukocyte receptor with distinct roles in human cancers (Review). Int J Oncol. 2013;43:1343–50.
69. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. Nat Biotechnol. 2020;38:970–9.
70. Nowak I, Malinowski A, Barcz E, Wilczyński JR, Wagner M, Majorczyk E, et al. Possible role of HLA-G, LILRB1 and KIR2DL4 gene polymorphisms in spontaneous miscarriage. Arch Immunol Ther Exp (Warsz). 2016;64:505–14.
71. Dufour EM, Deroche A, Bae Y, Muller WA. CD99 is essential for leukocyte diapedesis in vivo. Cell Commun Adhes. 2008;15:351–63.
72. Albini A, Di Guardo G, Noonan DM, Lombardo M. The SARS-CoV-2 receptor, ACE-2, is expressed on many different cell types: implications for ACE-inhibitor- and angiotensin II receptor blocker-based cardiovascular therapies. Intern Emerg Med. 2020;15:759–66.
73. Corley MJ, Pang APS, Dody K, Mudd PA, Patterson BK, Seethamraju H, et al. Genome-wide DNA methylation profiling of peripheral blood reveals an epigenetic signature associated with severe COVID-19. J Leukoc Biol. 2021;110:21–6.
74. Balnis J, Madrid A, Hogan KJ, Drake LA, Chieng HC, Tiwari A, et al. Blood DNA methylation and COVID-19 outcomes. Clin Epigenetics. 2021;13:118.
75. Castro de Moura M, Davalos V, Planas-Serra L, Alvarez-Errico D, Arribas C, Ruiz M, et al. Epigenome-wide association study of COVID-19 severity with respiratory failure. EBioMedicine. 2021;66:103339.
76. Yao C, Bora SA, Parimon T, Zaman T, Friedman OA, Palatinus JA, et al. Cell-type-specific immune dysregulation in severely ill COVID-19 patients. Cell Rep. 2021;34:108590.
77. de la Rica L, Rodríguez-Ubreva J, García M, Islam ABMMK, Urquiza JM, Hernando H, et al. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol. 2013;14:1–21.
78. Xu G, Qi F, Li H, Yang Q, Wang H, Wang X, et al. The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing. Cell Discov. 2020;6:73.
79. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science. 2017;356:eaah4573.
80. Delano MJ, Ward PA. Sepsis-induced immune dysfunction: can immune therapies reduce mortality? J Clin Invest. 2016;126:23–31.
81. Venet F, Demaret J, Gossez M, Monneret G. Myeloid cells in sepsis-acquired immunodeficiency. Ann N Y Acad Sci. 2021;1499:3–17.
82. Boks MA, Kager-Groenland JR, Haasjes MSP, Zwaginga JJ, van Ham SM, ten Brinke A. IL-10-generated tolerogenic dendritic cells are optimal for functional regulatory T cell induction–a comparative study of human clinical-applicable DC. Clin Immunol. 2012;142:332–42.
83. Matsuyama T, Kubli SP, Yoshinaga SK, Pfeffer K, Mak TW. An aberrant STAT pathway is central to COVID-19. Cell Death Differ. 2020;27:3209–25.
84. Varshney B, Agnihothram S, Agnihotram S, Tan Y-J, Baric R, Lal SK. SARS coronavirus 3b accessory protein modulates transcriptional activity of RUNX1b. PLoS ONE. 2012;7:e29542.
85. Xu J, Lee MH, Chakhtoura M, Green BL, Kotredes KP, Chain RW, et al. STAT2 is required for TLR-induced murine dendritic cell activation and cross-presentation. J Immunol. 2016;197:326–36.
86. Vandewalle J, Libert C. Glucocorticoids in sepsis: to be or not to be. Front Immunol. 2020;11:1318.
87. Dantzer R. Neuroimmune interactions: from the brain to the immune system and vice versa. Physiol Rev. 2018;98:477–504.
88. Conti P, Ronconi G, Caraffa A, Gallenga C, Ross R, Frydas I, et al. Induction of pro-inflammatory cytokines (IL-1 and IL-6) and lung inflammation by Coronavirus-19 (COVI-19 or SARS-CoV-2): anti-inflammatory strategies. J Biol Regul Homeost Agents. 2020;34:327–31.
89. Belfer I, Buzas B, Hipp H, Dean M, Evans C, Lorincz I, et al. Haplotype structure of inflammatory cytokines genes (IL1B, IL6 and TNF/LTA) in US Caucasians and African Americans. Genes Immun. 2004;5:505–12.
90. Adcock IM, Nasuhara Y, Stevens DA, Barnes PJ. Ligand-induced differentiation of glucocorticoid receptor (GR) trans-repression and transactivation: preferential targetting of NF-kappaB and lack of I-kappaB involvement. Br J Pharmacol. 1999;127:1003–11.
91. Cain DW, Cidlowski JA. After 62 years of regulating immunity, dexamethasone meets COVID-19. Nat Rev Immunol. 2020;20:587–8.
92. Vabret N, Britton GJ, Gruber C, Hegde S, Kim J, Kuksin M, et al. Immunology of COVID-19: current state of the science. Immunity. 2020;52:910–41.
93. Dress RJ, Ginhoux F. Monocytes and macrophages in severe COVID-19 - friend, foe or both? Immunol Cell Biol. 2021;99:561–4.
94. Szabo PA, Dogra P, Gray JI, Wells SB, Connors TJ, Weisberg SP, et al. Longitudinal profiling of respiratory and systemic immune responses reveals myeloid cell-driven lung inflammation in severe COVID-19. Immunity. 2021;54:797-814.e6.
95. Schultze JL, Mass E, Schlitzer A. Emerging principles in myelopoiesis at homeostasis and during infection and inflammation. Immunity. 2019;50:288–301.
96. Mann ER, Menon M, Knight SB, Konkel JE, Jagger C, Shaw TN, et al. Longitudinal immune profiling reveals key myeloid signatures associated with COVID-19. Sci Immunol. 2020;5:eabd6197.
97. Ehrlich M, Lacey M. DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. Epigenomics. 2013;5:553–68.
98. Li L, Gao Y, Wu Q, Cheng ASL, Yip KY. New guidelines for DNA methylome studies regarding 5-hydroxymethylcytosine for understanding transcriptional regulation. Genome Res. 2019;29:543–53.

99. Salman AA, Waheed MH, Ali-Abdulsahib AA, Atwan ZW. Low type I interferon response in COVID-19 patients: interferon response may be a potential treatment for COVID-19. Biomed reports. 2021;14:43.
100. Kondělková K, Vokurková D, Krejsek J, Borská L, Fiala Z, Ctirad A. Regulatory T cells (TREG) and their roles in immune system with respect to immunopathological disorders. Acta medica (Hradec Kral). 2010;53:73–7.
101. Single-cell multi-omics analysis of COVID-19 patients with pre-existing autoimmune diseases.Dataset. European Genome-Phenome Archive. 2022. Available from: https://ega-archive.org/datasets/EGAD00001007982.

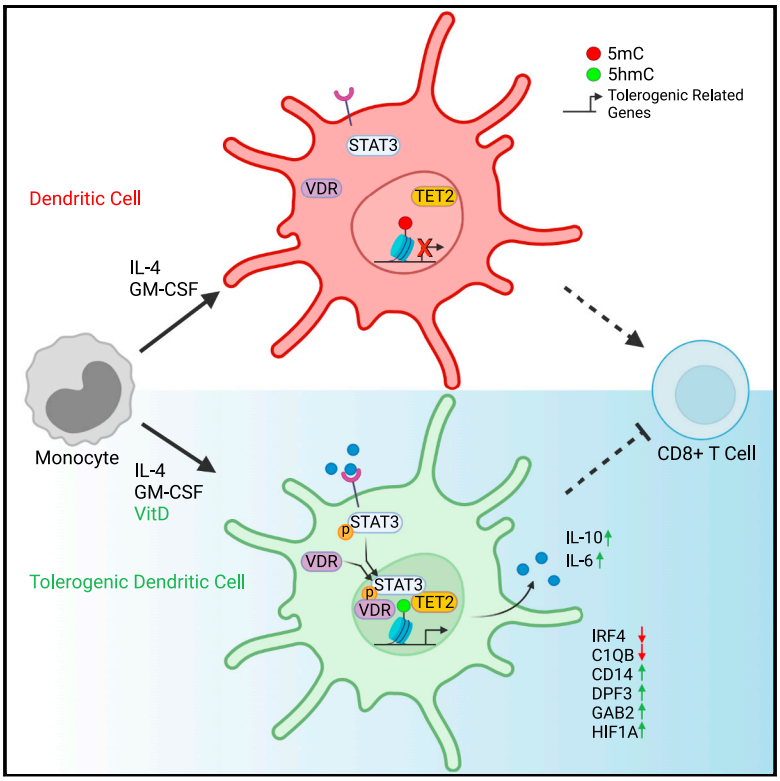
Publisher's Note

Cell Reports

Article

Vitamin D receptor, STAT3, and TET2 cooperate to establish tolerogenesis

Graphical abstract



Highlights

- Vitamin D induces DNA demethylation at VDR binding sites in dendritic cells (DCs)
- Differentiation to tolerogenic DCs associates with IL-6-JAK-STAT3 pathway activation
- VDR, STAT3, and TET2 interact with each other in tolerogenic DCs
- Pharmacological inhibition of JAK2 reverts vitamin D-induced DC tolerogenesis

Authors

Francesc Català-Moll, Anna G. Ferreté-Bonastre, Gerard Godoy-Tena, ..., Javier Rodríguez-Ubreva, Tianlu Li, Esteban Ballestar

Correspondence

eballestar@carrerasresearch.org

In brief

Català-Moll et al. show that vitamin D induces DNA demethylation and transcriptional activation at VDR binding sites, as well as IL-6-JAK-STAT3 pathway activation associated with acquisition of tolerogenesis by dendritic cells. VDR, STAT3, and TET2 interact with each other. Pharmacological inhibition of JAK2 reverts vitamin D-induced tolerogenic properties of DCs.

Català-Moll et al., 2022, Cell Reports *38*, 110244
January 18, 2022
https://doi.org/10.1016/j.celrep.2021.110244

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Vitamin D receptor, STAT3, and TET2 cooperate to establish tolerogenesis

Francesc Català-Moll,[1,4] Anna G. Ferreté-Bonastre,[1,4] Gerard Godoy-Tena,[1,4] Octavio Morante-Palacios,[1] Laura Ciudad,[1] Laura Barberà,[1] Federico Fondelli,[2,3] Eva M. Martínez-Cáceres,[2,3] Javier Rodríguez-Ubreva,[1] Tianlu Li,[1,4] and Esteban Ballestar[1,5,*]

[1]Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), Badalona, 08916 Barcelona, Spain
[2]Division of Immunology, Germans Trias i Pujol Hospital, LCMN, Germans Trias i Pujol Research Institute (IGTP), Badalona, 08916 Barcelona, Spain
[3]Department of Cell Biology, Physiology, Immunology, Autonomous University of Barcelona, Bellaterra, 08193 Barcelona, Spain
[4]These authors contributed equally
[5]Lead contact
*Correspondence: eballestar@carrerasresearch.org
https://doi.org/10.1016/j.celrep.2021.110244

SUMMARY

The active form of vitamin D, 1,25-dihydroxyvitamin D3, induces a stable tolerogenic phenotype in dendritic cells (DCs). This process involves the vitamin D receptor (VDR), which translocates to the nucleus, binds its cognate genomic sites, and promotes epigenetic and transcriptional remodeling. In this study, we report the occurrence of vitamin D-specific DNA demethylation and transcriptional activation at VDR binding sites associated with the acquisition of tolerogenesis *in vitro*. Differentiation to tolerogenic DCs associates with activation of the IL-6-JAK-STAT3 pathway. We show that JAK2-mediated STAT3 phosphorylation is specific to vitamin D stimulation. VDR and the phosphorylated form of STAT3 interact with each other to form a complex with methylcytosine dioxygenase TET2. Most importantly, pharmacological inhibition of JAK2 reverts vitamin D-induced tolerogenic properties of DCs. This interplay among VDR, STAT3, and TET2 opens up possibilities for modulating DC immunogenic properties in clinics.

INTRODUCTION

Dendritic cells (DCs) are a heterogeneous group of innate immune cells that have a key role in initiating adaptive responses. Also, DCs are not only central for coordinating immune responses against a threat but also needed to regulate the immune system at steady state and for inducing immune tolerance (Morante-Palacios et al., 2021). Like in other myeloid cell populations, the immunological properties of DCs vary with the environment. In general, terminal myeloid cell differentiation is highly dependent on the activation of specific signaling pathways in response to extracellular signals, such as inflammatory cytokines, hormones, vitamins, and other factors (Álvarez-Errico et al., 2015), which determine the immunogenicity of the resulting myeloid cells. The activation of signaling pathways leads to the activation of specific sets of transcription factors (TFs). Sequence-specific DNA binding of TFs is a pivotal process for establishing gene expression patterns in concert with the epigenetic machinery that determines cell identity and function (Monticelli and Natoli, 2017). Recent evidence has shown that several TFs are associated with DNA demethylation to increase genomic accessibility of their binding genomic regions, thus facilitating the binding of subsequent TFs (Mahé et al., 2017). In this regard, methylcytosine dioxygenase ten-eleven translocation (TET2), the most relevant enzyme involved in active DNA demethylation in the myeloid compartment, can interact with a variety of TFs, such as PU.1, C/EBPα, KLF4, and others, in order to facilitate their recruitment to different genomic regions (Costa et al., 2013; Guilhamon et al., 2013; de la Rica et al., 2013; Lio et al., 2016; Mendes et al., 2021; Sardina et al., 2018; Wang et al., 2015; Xiong et al., 2016). Recently, it has been demonstrated that TET2 mutations, which are frequent in myeloid leukemias, lead to DNA hypermethylation of enhancer regions and changes in the subsequent binding of TFs, particularly members of the basic helix-loop-helix (bHLH) TF family (Rasmussen et al., 2019). This suggests that TET2 recruitment by TFs leads to epigenetic remodeling that facilitates the binding of subsequent TFs (Rasmussen et al., 2019). Moreover, a reciprocal relationship between DNA methylation and histone modifications has long been established. TET2 has been not only described to modulate trimethylation of K4 of histone H3 (H3K4me3) (Deplus et al., 2013), a mark of active transcription, but also shown to coordinate trimethylation of K27 of histone H3 (H3K27me3), a mark of heterochromatin, in an inverse manner (Ichiyama et al., 2015).

Calcitriol (1,25-dihydroxyvitamin D3), the active form of vitamin D3 (henceforth referred to as vitamin D), is a major modulator of the immune system (Barragan et al., 2015; Carlberg, 2019; Mora et al., 2008). DCs are the most susceptible cell type to vitamin D in a mixed immune population (Mora et al., 2008). In these cells, vitamin D can generate a stable maturation-resistant tolerogenic

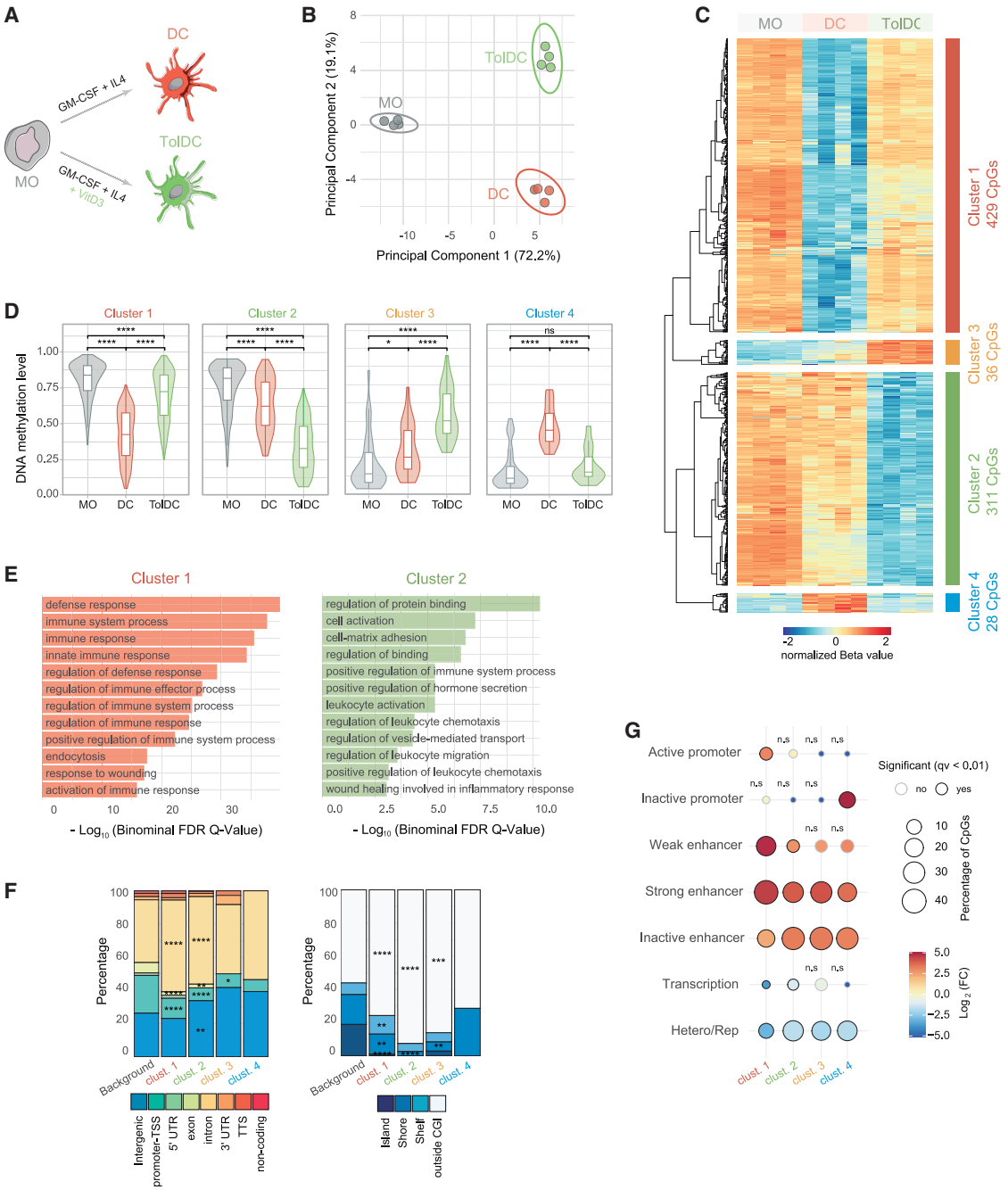


Figure 1. DNA methylation dynamics throughout vitamin D-exposed dendritic-cell differentiation (n = 4, two independent experiments)

(A) Schematic overview of the differentiation model from human peripheral blood MOs to DCs and TolDCs.

(B) Principal-component analysis of differentially methylated CpGs.

(C) DNA methylation heatmap and cluster analysis of four paired samples of MOs and their derived DCs and TolDCs at day 5 of differentiation. The heatmap includes all CpG-containing probes displaying significant methylation changes (differential beta value ≥ 0.2 and q value < 0.05) only in the TolDC-DC comparison. The color annotation of the lateral bar represents the membership to cluster 1 in red (DC-specific DNA demethylation), cluster 2 in green (TolDC-specific DNA demethylation), cluster 3 in orange (TolDC-specific DNA hypermethylation), and cluster 4 in blue (DC-specific DNA hypermethylation).

(D) Box and violin plots summarizing the distribution of DNA methylation levels per cell type and cluster.

(E) Gene ontology (GO) terms associated with CpGs from cluster 1 (red) and cluster 2 (green) as analyzed by GREAT software. Bars represent log-transformed binomial q values of the GO term enrichment.

(F) Location proportions of CpGs from each cluster in the context of CpG islands (CGIs) (right) and gene-related regions (left).

(legend continued on next page)

phenotype *in vitro*, with a low level of expression of immunogenic molecules, such as HLA-DR, CD80, and CD86, and increased interleukin (IL)-10/IL-12p70 ratios that are maintained even after removal of the compound (Van Halteren et al., 2002). After ligand recognition, vitamin D receptor (VDR) translocates to the nucleus and acts not only as a TF, controlling the expression of a set of immune and metabolic genes (Carlberg, 2019; Ferreira et al., 2013), but also as a repressor of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB) at different levels (Carlberg, 2019; Fetahu et al., 2014). Several studies have shown the capacity of VDR to interact with a range of TFs, including PU.1 and GABPA, and with chromatin remodeling and histone modification enzymes, such as BRD7 and KDM6B (Pereira et al., 2011; Seuter et al., 2017, 2018; Wei et al., 2018). Previous work has shown that vitamin D may induce DNA methylation changes in myeloid cells (Ong et al., 2021). However, the molecular mechanism that leads to the acquisition of differential methylation patterns remains unexplored.

Vitamin D supplementation is generally used as a preventive agent or a co-adjuvant for diseases with underlying autoimmune or pro-inflammatory states (Bscheider and Butcher, 2016; Dankers et al., 2017). DCs represent an excellent target of vitamin D to dampen autoimmunity and inflammation, not only because these myeloid cells express the whole set of enzymes to generate the active form of vitamin D (Mora et al., 2008) but also because of their unique role as initiators of immune responses. However, the role of DCs in vitamin D-mediated immunomodulation is not fully understood. In addition, DCs with tolerogenic function (TolDCs) have become a promising immunotherapeutic tool for reinstating immune tolerance in autoimmune diseases and in allogeneic bone marrow and solid organ transplantation (Morante-Palacios et al., 2021). The stability of the tolerogenic phenotype suggests that regulatory mechanisms that allow the maintenance of stable changes of gene expression are involved. In this sense, DNA methylation is a major epigenetic modification closely involved in the acquisition or stabilization of transcriptional states (Luo et al., 2018). Peripheral blood monocyte (MO)-derived DCs represent a useful model for studying the properties of DCs. It has been previously described that DCs differentiated from isolated MOs by the addition of granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-4 *in vitro* closely resemble CD1c$^+$ DCs at the transcriptional level (Goudot et al., 2017). Exposure of MO-derived DCs to vitamin D results in the inhibition of differentiation and maturation into potent antigen-presenting cells and gain in the capacity to inhibit T cell proliferation (Piemonti et al., 2000). Similarly, CD1c$^+$ DCs cultured *in vitro* with vitamin D for 2 days acquire a typical semi-mature phenotype after exposure to a DC maturation cocktail, with low CD83 expression, and a tolerogenic phenotype, as they suppressed alloimmunity *in vivo*, in a mouse model (Chu et al., 2012).

In this study, we studied epigenetic determinants critical for the acquisition of tolerogenic properties during *in vitro* human MO-derived DC differentiation in the presence of vitamin D. We demonstrate an interplay between VDR and the Janus kinase (JAK) 2/signal transducer and activator of transcription (STAT) 3 pathway associated with the generation of a specific TET-dependent DNA demethylation signature in TolDCs. It involves a direct physical interaction between VDR, STAT3, and TET2 that leads to the acquisition and stabilization of the tolerogenic properties of DCs in the presence of vitamin D.

RESULTS

Vitamin D induces the acquisition of a specific DNA methylation profile associated with tolerogenesis during *in vitro* DC differentiation

To investigate the effects of vitamin D in DNA methylation during the acquisition of tolerogenic properties by DCs, we first differentiated *in vitro* peripheral blood MOs from human donors to DCs and TolDCs for 6 days using GM-CSF and IL-4 in the absence and presence of vitamin D, respectively (Figure 1A). As previously described (Penna and Adorini, 2000; Piemonti et al., 2000), TolDCs had higher levels of the surface markers CD14 and CD11b and lower levels of HLA-DR, CD1a, and CD86 than did DCs (Figure S1A). To confirm the resemblance between our *in vitro* model with *in vivo* DCs, we integrated the expression profiles of MOs, DCs (12 h and 120 h), and TolDCs (12 h and 120 h) (Széles et al., 2009) with previously published expression datasets (Goudot et al., 2017; Segura et al., 2013) from MOs, *in vitro*-derived DCs and macrophages (MACs), and *in vivo* DCs and MACs. According to t-distributed stochastic neighbor embedding (t-SNE) analysis, TolDCs (differentiated in the presence of vitamin D), among different DC subsets, are the ones nearer different MAC types with immunosuppressive phenotypes (Figure S1B).

In concordance with previous studies (Piemonti et al., 2000), we observed that TolDCs were able to inhibit CD8$^+$ T cell proliferation *in vitro*, in contrast to DCs, confirming their immunosuppressive properties (Figure S1C). Furthermore, we also observed increased levels of VDR in the nucleus following vitamin D exposure, in agreement with previous studies, suggesting that VDR preferentially acts in the nucleus (Figure S1D). Altogether, our results confirmed the validity of this *in vitro* model to generate and study TolDCs by the involvement of VDR through vitamin D exposure.

We then obtained and compared the DNA methylation profiles of MOs, DCs, and TolDCs using BeadChip arrays (see STAR Methods), which interrogate the methylation status of >850,000 CpG positions across the entire genome, covering 99% of the reference sequence genes. Principal-component analysis (PCA) showed that most of the variability observed at the DNA methylation level may be explained by events common to the two differentiation processes (principal component 1;

(G) Bubble chart depicting the enrichment (red) or depletion (blue) of the CpGs from each cluster in the chromatin states from DCs (Pacis et al., 2015). The circle filling color represents the logarithmic value of the ratio between the percentage of CpGs with the feature in each cluster and the percentage of CpGs with the feature in the background. Circle size indicates the percentage of CpGs from each cluster in the chromatin state, and the circle edge indicates the statistical significance of the enrichment (black: significant; no edge: not significant; q value < 0.01).
Statistical tests: paired two-tailed t test (D), Pearson correlation (E), and two-tailed Fisher's exact test (F and G) (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, ns = not significant). FDR, false discovery rate.

Figure 1B). However, the second principal component is capable of clustering DCs and TolDCs separately (Figure 1B). Differentiation mainly resulted in DNA demethylation in which there were both condition-specific demethylation events and demethylation events common to both differentiation processes (Figure S1E). A small proportion of DNA methylation changes was attributed to gains of DNA methylation during differentiation (Figure S1E). Hierarchical clustering of differentially methylated CpGs between DCs and TolDCs (adjusted $p < 0.05$ and absolute differential beta ≥ 0.2) revealed four main groups of CpG sites (Figures 1C and 1D and Table S1): a group of CpGs that underwent specific demethylation in DCs (cluster 1: 429 CpGs); a second group that was specifically demethylated in TolDCs (cluster 2: 311 CpGs); another group that gained methylation in TolDCs (cluster 3: 36 CpGs); and finally a group of CpGs with DC-specific gains in DNA methylation (cluster 4: 28 CpGs).

To confirm these observations in the context of *in vivo* circulating DCs, we obtained the DNA methylation profiles of whole blood-isolated cDCs ($CD1c^+$ DCs) cultured in the absence and presence of vitamin D for 3 days and observed that, similar to TolDCs generated *in vitro*, cDCs exposed to vitamin D underwent DNA demethylation in cluster 2 CpGs (Figure S1F). This confirmed that demethylation observed in cluster 2 CpGs were specific to vitamin D exposure.

Functional gene ontology (GO) analysis revealed that CpGs in cluster 1 are associated with immunological categories, such as defense and immune response, whereas those in cluster 2 are more highly enriched in cell activation, positive regulation of immune system process, and wound healing involved in inflammatory response (Figure 1E). For clusters 3 and 4, GO analysis did not show enrichment in any functional categories, probably due to their small size. In all clusters, the majority of changes occurred in introns and intergenic regions with underrepresentation of promoter-transcriptional start sites (TSSs). However, whereas cluster 1 exhibited a marked enrichment of intronic regions with respect to background, the other clusters were enriched in both intronic and intergenic locations (Figure 1F, left). Concordantly, CpGs of all clusters were observed to be located outside of CpG islands, particularly for cluster 2 (Figure 1F, right). Next, we mapped the chromatin states of the CpG sites undergoing changes in methylation in the four clusters using chromatin segmentation data generated in DCs (Pacis et al., 2015) (Figure 1G). We observed an enrichment in enhancer regions for all clusters and an enrichment for inactive promoters for cluster 4. Moreover, cluster 1 (DC-specific demethylation) was enriched in weak (H3K27ac + H3K4me1 + H3K4me3) and strong (H3K27ac + H3K4me1) enhancers, while cluster 2 (TolDC-specific demethylation) was more enriched in inactive enhancers (H3K4me1) in DCs, suggesting that these inactive regions in DCs are activated in TolDCs. In all, our results indicated that vitamin D-driven demethylation events occurred in regions that may play important roles in regulating gene expression and establishing the tolerogenic phenotype of TolDCs.

DNA demethylation in TolDCs is an active process and is associated with changes in gene expression

DNA methylation has long been established to influence gene expression (Jones, 2012), although the dynamics are complex and highly dependent on genomic location. CpGs that underwent TolDC-specific DNA demethylation during differentiation were largely situated in open seas corresponding to enhancers; hence, it is plausible to envision that they control gene expression, which results in the final tolerogenic phenotype. We therefore integrated our DNA methylation dataset with publicly available expression data generated in the same *in vitro* models (Széles et al., 2009). We observed a significant inverse relationship between levels of DNA methylation and mRNA expression at 12 h ($r = -0.5926$; $p = 4.90e\text{-}14$) and 5 days of differentiation ($r = -0.4108$; $p = 4.57e\text{-}11$) (Figure 2A). Furthermore, dividing cluster 1 and 2 CpGs based on their genomic location in relation to previously identified enhancer regions (Pacis et al., 2015), we observed that genes associated with cluster 1 CpGs located at active enhancers of DCs displayed higher expression levels in DCs than in TolDCs (Figure 2B).

To explore the dynamics of the relationship between DC- (cluster 1) and TolDC-specific demethylation (cluster 2), we performed bisulfite pyrosequencing and qRT-PCR in a selected group of genes of a set of samples over time. A few genes from each cluster were selected for further analysis based on the conditions that they had the maximum possible difference in DNA methylation during differentiation within their corresponding cluster, that they were differentially expressed, and that there were previous reports relating them with relevant immune properties. For instance, from cluster 1, we chose *IRF4* and *C1QB*, which are important for normal DC differentiation from MOs (Teh et al., 2011; Murphy et al., 2016), and from cluster 2, *CD14* and *DPF3* were selected for being specific markers of TolDCs and being involved in DC chemotaxis, respectively (Liu et al., 2019; Torres-Aguilar et al., 2010). Bisulfite pyrosequencing of these genes showed a high concordance ($r = 0.978$; $p < 2.2 \times 10^{-16}$) with the data obtained from the EPIC arrays (Figure 2C). DC-specific (cluster 1) genes, such as *IRF4* and *C1QB*, were upregulated in DCs in parallel with their specific DNA demethylation (Figure 2D). Similarly, for TolDC-specific (cluster 2) genes, such as *CD14* and *DPF3*, transcript upregulation occurred only in TolDCs in parallel with their corresponding DNA demethylation (Figure 2E). In agreement with previous reports, stimulus-induced DNA demethylation occurred succeeding specific gene expression changes (Pacis et al., 2019). In all, our results suggested that vitamin D-driven DNA demethylation occurred in association with upregulation of TolDC-specific genes.

To further characterize the mechanisms driving DNA demethylation during MO-to-DC and MO-to-TolDC differentiation, we next investigated whether the demethylation was due to active demethylation or replication-mediated passive demethylation. Utilizing BrdU proliferation assay, no proliferation was observed in DCs and TolDCs up to 6 days of differentiation (Figure S2A); hence, all DNA demethylation events observed were driven by active demethylation. In this regard, we and others have previously shown that loss of methylation in terminal differentiation from MOs is accompanied by a transient increase in 5-hydroxymethylcytosine (5hmC) and involves the participation of TET2 methylcytosine dioxygenase (Garcia-Gomez et al., 2017; Klug et al., 2013). We then determined the 5hmC levels of CpGs that became demethylated during DC and TolDC differentiation and observed that there was indeed a gain of 5hmC in these

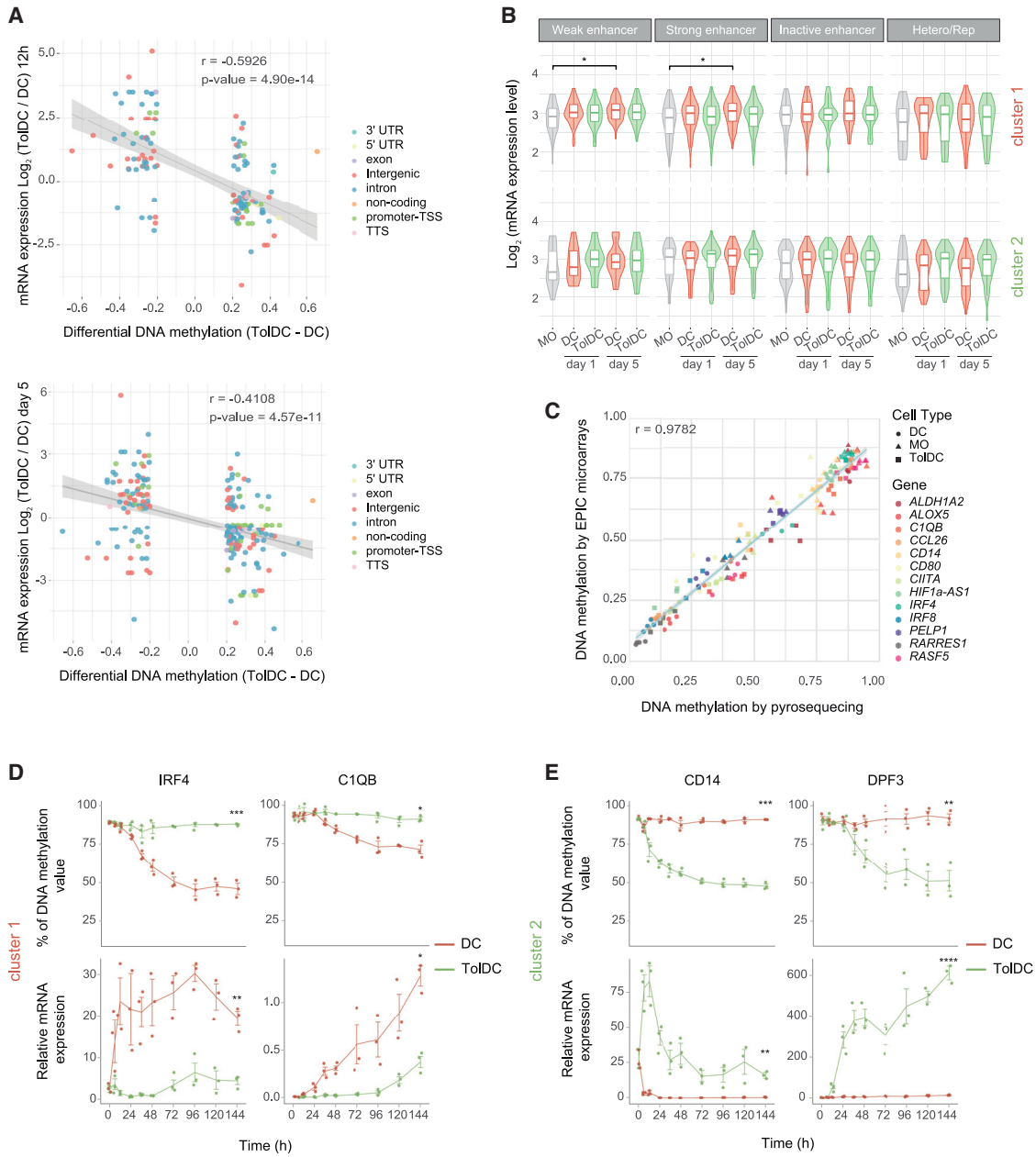


Figure 2. Integration of gene expression with DNA methylation

(A) Scatter plot showing the correlation between DNA methylation differences and gene expression changes between DCs and TolDCs at 12 h (top) and day 5 (bottom) of differentiation. Only differentially methylated CpGs are represented. Dot color indicates gene-related associations.

(B) Box and violin plots summarizing the mRNA expression levels per cell type of genes annotated to CpGs from cluster 1 (top) and cluster 2 (bottom) divided by chromatin state annotation of the associated CpG. Adjustment for multiple comparisons was performed with false discovery rate.

(C) Scatter plot showing the correlation between methylation array values and bisulfite pyrosequencing DNA methylation values (n = 4, two independent experiments).

(D) DNA methylation (top) and mRNA expression (bottom) kinetics of two representative examples of cluster 1 genes. CpGs studied include cg10630015 (IRF4) and cg04097715 (C1QB) (n = 3, one single experiment).

(E) DNA methylation (top) and mRNA expression (bottom) kinetics of two representative examples of cluster 2 genes (n = 3, two independent experiments). CpGs studied include cg05620710 (CD14) and cg25205844 (DPF3).

Statistical tests: Pearson correlation (A) and unpaired two-tailed t test (B, D, and E) (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001).

CpGs (Figure S2B). Finally, utilizing publicly available DNase-sequencing (seq) datasets from MOs (Feingold et al., 2004), we observed that more than 75% of cluster 2 CpGs corresponded to closed chromatin in MOs (Figure S2C), which reinforced the hypothesis that DNA demethylation was mediated by an active event. Altogether, our results suggested that specific active DNA demethylation following vitamin D exposure is mediated through methylcytosine dioxygenase activity, most likely associated with TET2.

VDR binding is associated with DNA demethylation and active chromatin during MO-to-TolDC differentiation

In concordance with previous work (Jakob et al., 1992), we observed that exposure to vitamin D during DC differentiation increased the nuclear levels of VDR (Figure S1D). Hence, it is plausible that liganded VDR plays a direct role in driving DNA demethylation following vitamin D exposure during TolDC differentiation. Hence, we performed chromatin immunoprecipitation (ChIP)-seq analysis of VDR in DCs and TolDCs. First, we observed that exposure to vitamin D during TolDC differentiation led to a sharp increase in overall VDR genomic binding (Figures 3A and 3B).

Interestingly, motif discovery analysis revealed promiscuity of VDR with respect to its genomic binding preferences, with only 37% of regions having the canonical VDR binding motif (Figure 3C), which suggests the cooperation of VDR with other TFs during TolDC differentiation. Second, functional annotation of VDR-bound genes revealed enrichment of immune- and signaling-related categories, such as myeloid and granulocyte activations and cytokine receptor activity (Figure 3D). In fact, several genes previously described to be related to the tolerogenic properties of TolDCs, such as *IL10*, *ANXA1*, and *CD163* (Navarro-Barriuso et al., 2018), are direct targets of VDR (Table S2). Third, global inspection of VDR genomic occupancy showed that VDR preferentially binds to promoters and introns in comparison with background (Figure 3E, left). We also observed enrichment of VDR binding in CpG islands, shores, and shelves, which was compatible with the enrichment noted in promoters (Figure 3E, right). Annotation of VDR peaks in relation to previously published data of DC chromatin states (Pacis et al., 2015) showed the preference of VDR for binding regions that correspond to promoters and enhancers in DCs (Figure 3F).

To further characterize the relationship between VDR and DNA methylation, we overlapped our generated DNA methylation data with VDR ChIP-seq data and observed a specific enrichment of VDR binding in TolDCs to CpGs that became demethylated in TolDCs (cluster 2), and this was not observed for the other clusters (Figures 4A, 4B, and S3A). In fact, we observed that over 40% of CpG sites in cluster 2 had significant VDR binding (Figure 4C). For instance, cluster 2 CpGs mapped to genes, such as *GAB2* and *HIF1A*, situated within the binding peaks of VDR in TolDCs (Figure 4D) and located in closed chromatin regions in MOs (Figure S3B). These genes are of particular interest because GAB2 has been implicated in phosphatidylinositol 3-kinase (PI3K) pathway activation (Pratt et al., 2000), a pathway implicated in DC tolerogenesis (Ferreira et al., 2015). Furthermore, hypoxia-inducible factor 1-alpha (HIF1A) is a key factor for the tolerogenic properties of myeloid-derived suppressor cells (MDSCs) in the tumor microenvironment (Corzo et al., 2010). The dynamics of DNA methylation and gene expression of these two genes confirmed specific DNA demethylation in TolDC, and differential gene expression changes in relation to DCs (Figures 4E and 4F).

As indicated in the introduction, TET-mediated demethylation is associated with histone modifications, such as H3K4me3 (Deplus et al., 2013) and H3K27me3 (Ichiyama et al., 2015). Hence, we speculated that changes in DNA methylation were accompanied by changes in histone modifications, and their dynamics might be associated with VDR recruitment following vitamin D exposure. Therefore, we performed ChIP-qPRC of VDR together with these activating (H3K4me3) and repressive (H3K27me3) histone modifications. We also added an antibody against H3 acetylation (H3ac), characteristic of active chromatin. To discriminate between the effects of a tolerogenic phenotype acquired through a 6-day differentiation and the effects directly caused by the presence of vitamin D in the medium, we performed ChIPs in MOs, DCs, TolDCs, and also DCs treated with vitamin D for 30 min (DC + vitD). First, we observed a significant increase in VDR binding (Figure 4G) in DCs treated with vitamin D and in TolDCs. Second, in the aforementioned cluster 2 genes *GAB2* and *HIF1A*, we only observed a significant increase, associated with VDR binding, for H3ac (Figure 4G). This finding was extendable to other cluster 2 genes, such as *HOPX*, *IL6*, *INHBA*, and *LYRM1* (Figure S3C).

Hence, altogether, our data suggested the coordination between VDR binding, specific DNA demethylation, changes in histone H3 acetylation, and gene expression upregulation in TolDC differentiation.

Differentiation to DCs in the presence of vitamin D associates with activation of IL-6-JAK-STAT3 signaling pathway, and both VDR and STAT3 interact with TET2

Vitamin D, through its receptor VDR, induces changes in cytokine production and a profound metabolic reprogramming in human DC (Ferreira et al., 2015). For this reason, we hypothesized that autocrine/paracrine activation of secondary signaling pathways during differentiation could lead to the activation of a set of TFs downstream to VDR that could be relevant to TolDC differentiation. To explore this possibility, we adapted a tool initially designed to explore intercellular communication in bulk and single-cell expression data to test autocrine/paracrine signal activation (Browaeys et al., 2020). Note that our differentiation model does not allow to distinguish between autocrine or paracrine activation. With this approach, and using genes associated with both demethylation clusters with significant expression differences (fold-change <0.5 or >2, and adjusted $p < 0.05$) as input, we inferred potential ligands that may regulate these processes (Figure 5A). One of the most interesting ligands due to its role in immune suppression in the context of tumorigenesis is IL-6 (Park et al., 2017). In fact, the *IL6* gene is significantly overexpressed in TolDCs compared with DCs (Figure 5B), and its target genes were also observed to be overexpressed in TolDCs (Figure 5C).

We then performed gene set enrichment analysis (GSEA) of differentially expressed genes between DCs and TolDCs and visualized that genes differentially overexpressed in TolDCs were enriched in IL-6-JAK-STAT3 signaling pathway (Figure 5D).

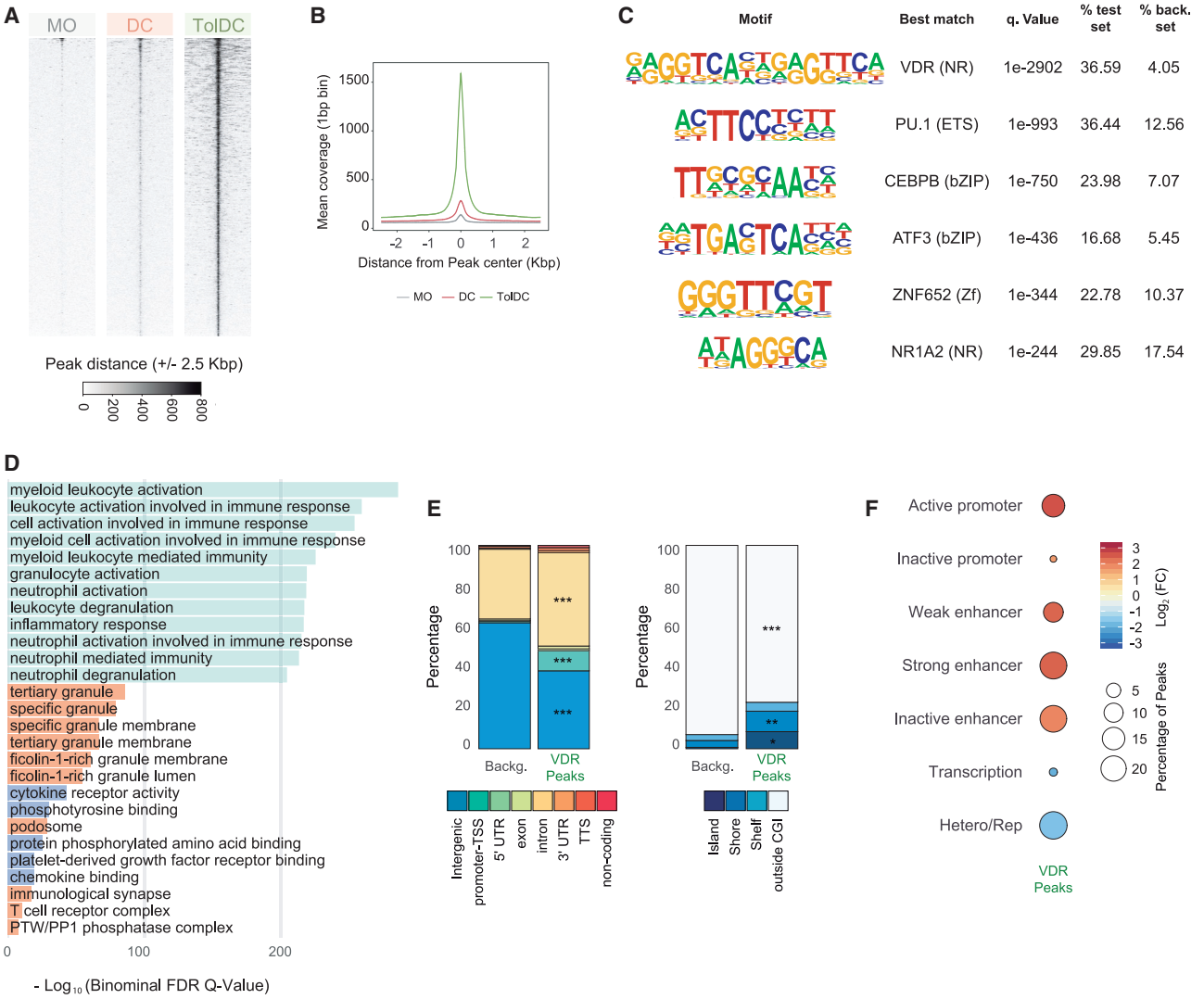


Figure 3. Genomic occupancy of vitamin D receptor (ChIP-seq, n = 2, one experiment)

(A) Heatmaps showing signal intensity of vitamin D receptor (VDR) ChIP-seq at ± 2.5 Kbp window of significant VDR peaks in MO, DCs and TolDCs (q value < 0.01 and irreproducible discovery rate [IDR] < 0.05).

(B) Composite plots of VDR ChIP-seq distribution ±2.5 Kbp around CpGs in MO (gray), DCs (red), and TolDCs (green) for significant VDR peaks. The statistics were computed by comparing the intensity averages of the entire window.

(C) Motif discovery analysis using HOMER software showing q values and the percentage of test and background regions with each motif.

(D) Results of gene set enrichment analysis using GREAT software. The plot depicts the top enriched terms for biological processes (green), molecular function (orange), and cellular component (purple) categories, based on adjusted p values from the binomial distribution.

(E) Location proportions of VDR peaks in the context of CpG islands (CGIs) (right) and gene-related regions (left).

(F) Bubble chart depicting the enrichment (red) or depletion (blue) of VDR peaks in the chromatin states of dendritic cells (Pacis et al., 2015). The circle filling represents the logarithm of the ratio between the percentage of VDR peaks with the feature and the percentage of the feature within the background. Circle size indicates the percentage of VDR peaks in the chromatin state, and the circle edge indicates the statistical significance of the enrichment (black: significant; no edge: not significant; q value < 0.01).

Statistical tests: two-tailed t test (A and B), cumulative binomial distribution (C and D), and two-tailed Fisher's exact test (E and F) (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001).

In fact, VDR binds in several regions upstream of the *IL6* gene TSS, suggesting that VDR directly regulates its expression (Figure 5E). Furthermore, we detected an increase in IL-6 production and release into the medium in TolDCs (Figure S4A), which was concordant with an upregulation of its gene expression compared with DCs (Figure 5B). Additionally, significant DNA demethylation was observed in 2 CpG sites of the promoter region of *IL6* in TolDCs, and this was coupled with a gain in 5hmC (Figures S4B and S4C), which suggested the involvement of TET2 in its regulation.

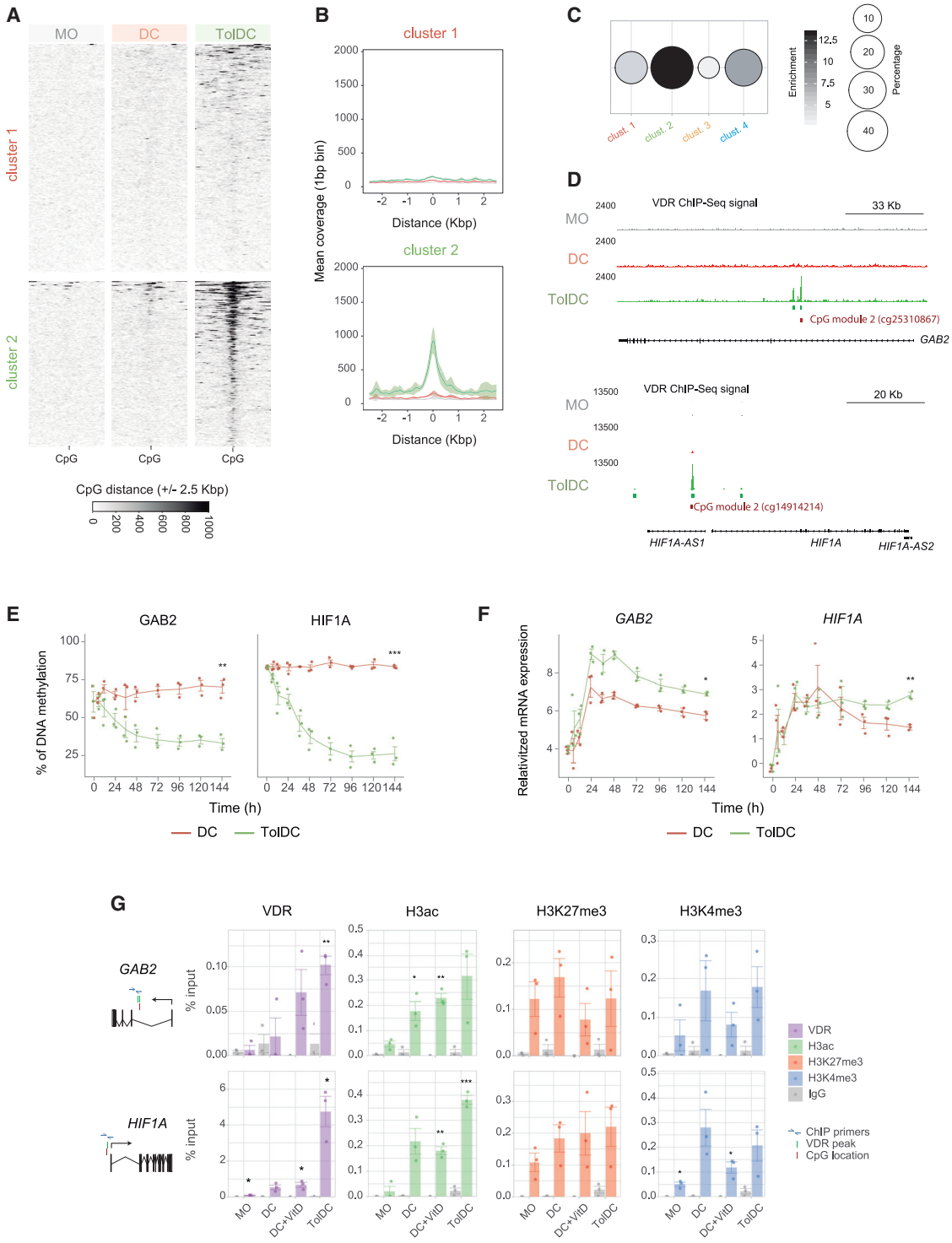


Figure 4. Binding of vitamin D receptor correlates with TolDC-specific DNA demethylation (ChIP-seq, n = 2, one single experiment; DNA methylation, n = 4, two independent experiments)

(A) Heatmaps showing signal intensity of vitamin D receptor (VDR) ChIP-seq at ± 2.5 Kbp window from CpGs of cluster 1 (top) and cluster 2 (bottom) in MO, DCs, and TolDCs.

(B) Composite plots of VDR ChIP-seq distribution ±2.5 Kbp around CpGs from cluster 1 (top) and cluster 2 (bottom) in MO (gray), DC (red), and TolDC (green). Smooth represents the CIs.

(legend continued on next page)

Remarkably, when we blocked IL-6 with an anti-IL-6 antibody during TolDC differentiation, we observed the production of decreased levels of IL-10 (Figure S4D), which is involved in tolerogenesis (Morante-Palacios et al., 2021). This result is consistent with recent findings in T helper type 1 (Th1) cells (Chauss et al., 2022). However, blocking IL-6 during TolDC differentiation did not result in a reduced ability of TolDCs to suppress $CD8^+$ T cell proliferation (Figure S4E). In contrast, in proliferation assays performed with TolDCs in the presence of anti-IL-6 antibody, we found slightly reduced suppression (Figure S4E). These results suggest that IL-6 is a contributor to the ability to suppress $CD8^+$ T cell proliferation by TolDCs but not critical to the acquisition of such properties during TolDC differentiation.

In parallel, we utilized DoRothEA (discriminant regulon expression analysis), a manually curated human regulon for estimating single-sample TF activities through the expression of their target genes (Garcia-Alonso et al., 2019) to analyze TF activities of several STATs in genes differentially expressed in TolDCs compared with DCs, and observed a specific increase in STAT3 activity that was not observed for other members of the STAT family, at 5 days of differentiation (Figure 5F). Furthermore, we observed a marked increase in phosphorylation of STAT3 in TolDCs compared with DCs, which was not observed for STAT5 (Figures 5G and S4E). Although a statistically significant increase in phosphorylation was observed for STAT1, this increase was not to the same extent as STAT3 and may be due to indirect activation, as previously described (Haan et al., 2005) (Figures 5G and S4F). Thus, our results suggested that vitamin D played a role in STAT3 activation.

To explore the possibility that the observed interplay between VDR and STAT3 involves a physical interaction, we performed co-immunoprecipitation experiments in TolDCs. Our analysis revealed a specific interaction between VDR and phosphorylated (p)-STAT3 in TolDCs (Figure 5H). We also observed that both VDR and p-STAT3 interacted with TET2 (Figure 5I), which suggests that these two TFs play a role in the targeting of TET2-mediated demethylation to their cognate sites.

Inhibition of JAK2-mediated STAT3 activation affects the acquisition of vitamin D-dependent tolerogenesis

We investigated the consequences of inhibiting the JAK2-STAT3 pathway by using TG101348, a pharmacological inhibitor of JAK2 (Lasho et al., 2008), during DC and vitamin D-dependent TolDC differentiation. Following TG101348 treatment, we confirmed the inhibition of STAT3 phosphorylation by western blot (Figure 6A). Given that TG101348 is an inhibitor of JAK2, and therefore can affect upstream signaling of STAT1, STAT3, and STAT5, we checked their phosphorylation and observed that the partial inhibition of p-STAT5 and p-STAT1 did not reach statistical significance in TolDCs, unlike p-STAT3 (Figures S5A and S5B). TG101348 treatment also resulted in a sharp decrease in the production of IL-10 (Figure 6B), an archetypical anti-inflammatory cytokine that is also a *bona fide* target for STAT3 (Schaefer et al., 2009; Ziegler-Heitbrock et al., 2003). In fact, IL-10 secretion by TolDCs is a contributor to the suppression of $CD8^+$ T cell proliferation that is halted when adding anti-IL-10 to proliferation assays (Figure S4E). We also tested the effects of JAK2 inhibition on surface markers and observed that JAK2 inhibition resulted in an increase of CD14 and CD86 protein levels and downregulation of CD1a and CD11b (Figure 6C). In parallel, we investigated the effects of JAK2 inhibition on the DNA methylation and expression levels of TolDC-specific demethylated genes. We did not observe any clear reversion of DNA demethylation (Figure 6D), but we did note alterations at the transcriptional level (Figure 6E). Changes were observed not only in cluster 2 genes (TolDC-specific), such as *CD14* and *DPF3*, but also in those of cluster 1, such as *IRF4* and *RASF5* (Figure 6E). These are likely to be the result of the partial inhibition of phosphorylation of STAT1 and STAT5, which might also be involved in activating these and other DC and TolDC genes.

Most importantly, JAK2 inhibition by TG101348 treatment during differentiation resulted in the loss of the ability to suppress $CD8^+$ T cell proliferation of DC differentiated in the presence of vitamin D. This reinforces the idea that the activities of VDR and the JAK2-STAT3 pathway coordinate the acquisition of tolerogenic properties of DCs in the presence of vitamin D (Figure 6F).

DISCUSSION

In this study, we demonstrate that vitamin D is able to induce tolerogenesis in DCs through a mechanism that involves VDR-specific demethylation and activation of key immune genes in a manner that is coordinated with JAK2-mediated STAT3 activation. VDR not only is able to orchestrate a direct response on key immune targets but also associates with activation of the IL-6-JAK-STAT signaling pathway. We also prove the recruitment of TET2 and p-STAT3 by VDR, associated with the

(C) Bubble plot representation of significant VDR binding enrichment in each cluster of CpGs. Dots are colored according to their enrichment value, defined as the logarithm of the ratio between the percentage of CpGs with VDR peak within the cluster and the percentage of CpGs with the VDR peak within the background. Bubble size corresponds to the percentage of CpGs in each cluster overlapping with significant VDR peaks. The presence of a black border indicates significant enrichment (q value < 0.01).
(D) VDR ChIP-seq signal profiles in the vicinity of the representative genes of CpGs from cluster 2. VDR signals are colored by cell type. At the bottom, the significant VDR binding sites are shown in green and CpG position in red.
(E) DNA methylation kinetics of two representative CpGs annotated to GAB2 (cg25310867) and HIF1A (cg14914214) in DCs and TolDCs (n = 3, one experiment).
(F) Gene expression kinetics of GAB2 and HIF1a in DCs and TolDCs (n = 3, one experiment).
(G) Bar plot representation of ChIP-qPCR results for VDR binding and three histone modifications (H3ac, H3K27me3, and H3K4me4) in the vicinity of VDR peaks close to GAB2 and HIF1A gene sequences (n = 3, one experiment). This analysis was performed in MOs, DCs, and TolDCs and DC + vitD. DC + vitD involves adding vitamin D for 30 min at the end of a 6-day differentiation to DCs. Immunoprecipitation with IgG was used as control. The location of the ChIP primers, the CpG site and the VDR peaks are indicated.
Statistical tests: two-tailed Fisher's exact test (C) and unpaired two-tailed t test (E, F, and G) (*p < 0.05; **p < 0.01; ***p < 0.001).

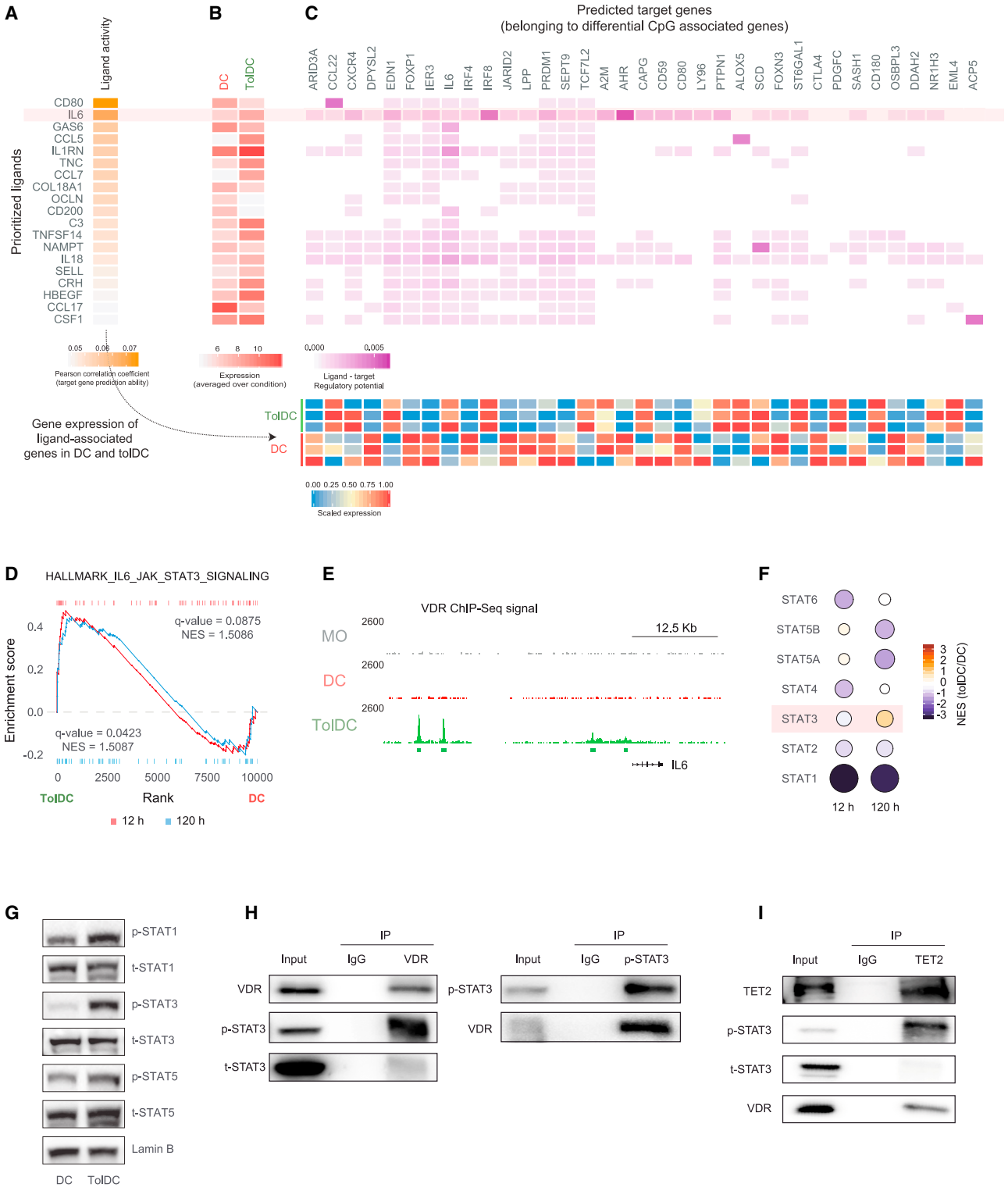


Figure 5. Vitamin D-dependent autocrine/paracrine activation of the IL-6-JAK2-STAT3 pathway

(A) Heatmap showing ligand activity prediction based on the Pearson correlation with its target genes.

(B) Heatmap displaying average gene expression of ligands for DCs and TolDCs on day 5.

(C) Heatmap showing the regulatory potential of each ligand on the target genes based on nichenetr package database (upper panel) and the expression levels of these target genes in each sample (lower panel).

(legend continued on next page)

demethylation and activation of target genes. The essential role of the JAK2-STAT3 pathway in the acquisition of tolerogenesis is demonstrated by the functional impact of the pharmacological inhibition of this pathway.

Our results show the direct role of VDR in guiding TET-mediated DNA demethylation to specific genomic sites during TolDC differentiation. We have shown that, in the presence of vitamin D, VDR levels are increased in the nucleus and that interaction with p-STAT3 and TET2 occurs, thereby promoting TolDC-specific demethylation. A recurrent question in the DNA methylation field is whether DNA methylation is causally involved in shaping gene expression profiles or if it passively reflects transcriptional states (Schübeler, 2015). Our own data support both possibilities, and some DNA methylation changes appear to be more likely to occur after a change in expression than others (Pacis et al., 2019). In our study, we present evidence that TET-mediated demethylation acts as a mechanism facilitating subsequent participation of other TFs, in this case STAT3. In fact, the absence of interference with DNA demethylation, while activation is impeded following pharmacological inhibition of STAT3 phosphorylation, suggests that VDR-dependent demethylation is necessary and precedes STAT3-mediated gene activation. This proposed mechanism was consistent with the alterations in TF activity reported in TET2 knockout mice (Rasmussen et al., 2019). TET2-associated functions may ensure the binding of some TFs, thereby contributing to enhancer-dependent activity and gene expression.

Our study identifies a crucial role for the JAK2-STAT3 pathway in the acquisition of tolerogenesis in innate immunity. The involvement of STAT3 is also relevant in the context of MDSCs, which are also characterized by their tolerogenic properties (Corzo et al., 2009; Kumar et al., 2016). We show that the pharmacological impairment of STAT3 phosphorylation, by inhibiting JAK2, directly results in the loss of the tolerogenic properties of TolDCs, which facilitate T-cell proliferation, demonstrating the essential role of this pathway for the tolerogenic phenotype. Our results raise the possibility that tolerogenic properties can be reverted, not only in the context of vitamin D but also in others. These findings could be clinically relevant both in the context of pathological situations where tolerogenic properties are not desired, like in the tumor microenvironment or in metastatic processes (reviewed in DeVito et al., 2019), as well as in those where they are intentionally pursued (reviewed in Cauwels and Tavernier, 2020), including their therapeutic use in the treatment of inflammatory conditions, such as rheumatoid arthritis and multiple sclerosis (Morante-Palacios et al., 2021).

Limitations of the study

One of the limitations of our current study is that we have not fully explored the impact of the VDR and the IL-6-JAK-STAT3 pathway *in vivo*, in patients treated or supplemented with vitamin D. It would have also been relevant to analyze the direct impact of STAT3 in the epigenetic remodeling in TolDCs, by analyzing their binding sites and associated expression changes. This partly limits our conclusions on the extent and relevance of STAT3 in determining the acquisition of the tolerogenic phenotype.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Differentiation of TolDCs and DCs from peripheral blood monocytes
  - Isolation and culture of peripheral blood DCs
- METHOD DETAILS
  - $CD8^+$ cell proliferation assay
  - BrdU proliferation assay
  - Flow cytometry
  - Cytokine measurements
  - Genomic DNA and total RNA extraction
  - Bisulfite (BS) and oxidative-bisulfite (oxBS) pyrosequencing
  - Real-time quantitative Polymerase Chain Reaction (RT-qPCR)
  - Western blot
  - Co-immunoprecipitation (Co-IP)
  - DNA methylation profiling
  - ChIP-seq analysis
  - ChIP-qPCR
  - Microarray reanalysis
  - Data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.110244.

(D) Gene set enrichment analysis of differentially expressed genes (fold-change < 0.5 or >2 and q-value < 0.05) at 12 h (red) and 120 h (blue). Results for the IL-6-JAK-STAT signaling pathway are shown.
(E) VDR ChIP-seq signal profiles in the vicinity of the *IL6* gene. VDR signals are colored by cell type. The significant VDR binding sites are shown below in green.
(F) Bubble chart depicting the TF activity predicted from mRNA expression of target genes with DoRothEA v2.0. The circle filling represents the normalized enrichment score (NES) (blue: more activity in DCs; red: more activity in TolDCs). Bubble size corresponds to the logarithm of adjusted p values.
(G) Representative western blot assays showing the phosphorylated and total protein levels of STAT1, STAT3, and STAT5 on day 3 of differentiation of DCs and TolDCs (n = 4, two independent experiments).
(H) Representative western blots showing the results of co-immunoprecipitation assays performed in MOs differentiated to DC and TolDC for 3 days. Protein extracts were immunoprecipitated using anti-VDR or anti-p-STAT3 antibodies (n = 3, two independent experiments).
(I) Representative western blots showing the results of co-immunoprecipitation assays performed in MOs differentiated to TolDC for 3 days (n = 3, two independent experiments). Protein extracts were immunoprecipitated using anti-TET2 antibodies.
In both (H and I), IgG was used as a negative control and total protein extract was used as input.

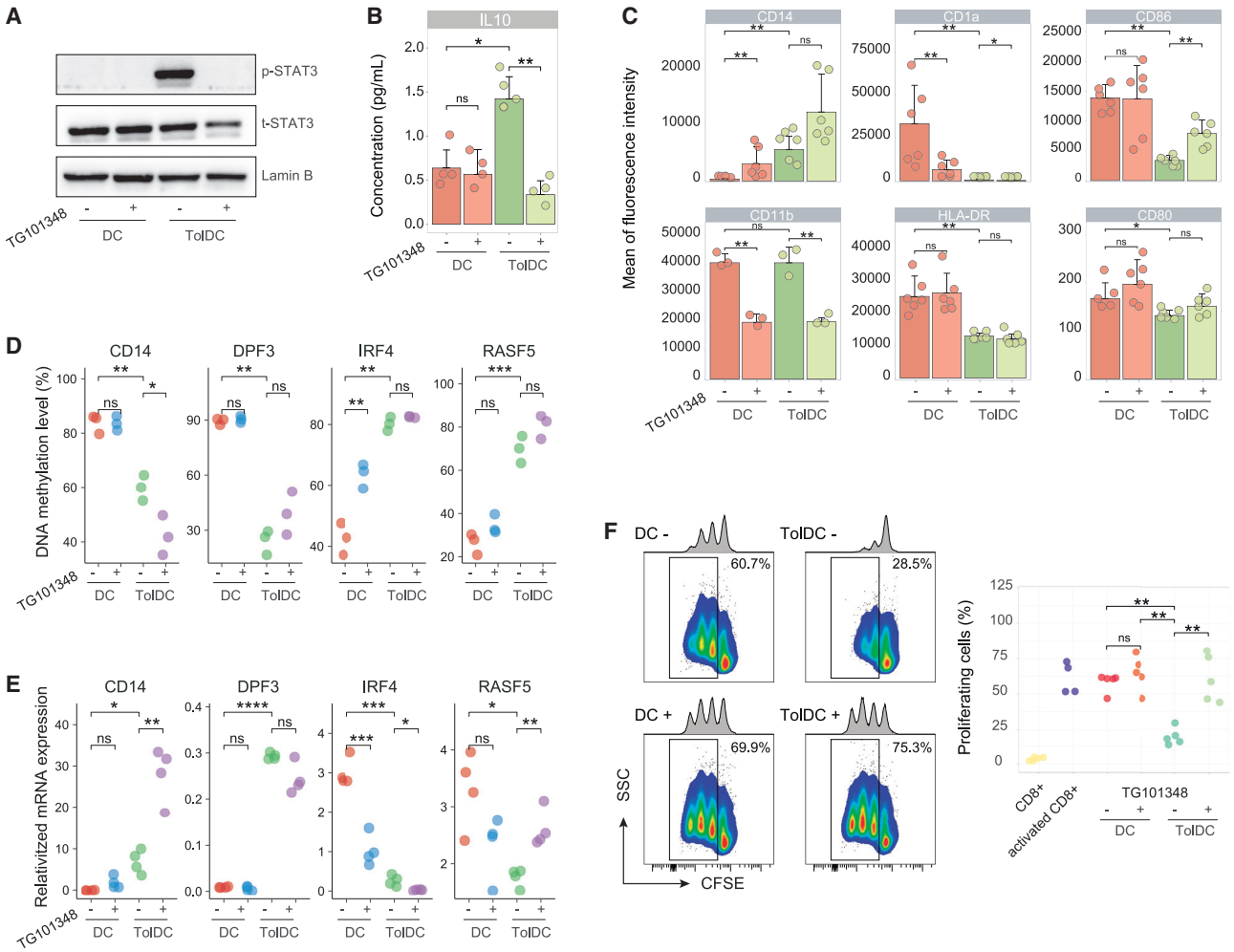


Figure 6. Inhibition of STAT3 phosphorylation reverts immunosuppressive properties of vitamin D exposed dendritic cells

(A) Representative western blot assays showing the effects of STAT3 at the protein phosphorylation level after pharmacological inhibition of JAK2 with TG101348 (n = 3, two experiments).

(B) Bar plot representation of the effect of JAK2 inhibition with TG101348 on IL-10 released by DCs and TolDCs (n = 4, two independent experiments). Protein levels were measured by ELISA.

(C) Bar plots showing the impact of JAK2 inhibition with TG101348 on membrane receptor expression (n = 6, two independent experiments). Protein levels were measured with flow cytometry.

(D) Dot plot representation of bisulfite pyrosequencing results of four example CpGs, two from cluster 1 (CD14 and DPF3) and two from cluster 2 (C1QB and RASF5), displaying the consequence of JAK2 inhibition with TG101348 in DCs and TolDCs (n = 3, two independent experiments).

(E) Dot plot showing mRNA expression of four example genes from cluster 1 and cluster 2 as measured by RT-qPCR, showing the effect of JAK2 inhibition with TG101348 in DCs and TolDCs (n = 4, two independent experiments). Expression was relativized with respect to RPL38 gene expression.

(F) Representative example and dot plot showing the effect on $CD8^+$ cell proliferation of DCs and TolDC generated from MO in presence or absence of TG101348 (n = 5, two independent experiments).

Statistical tests: two-tailed Wilcoxon rank-sum test (B–F) (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001).

ACKNOWLEDGMENTS

We are very grateful to Dr. José Luis Sardina for useful feedback. We thank CERCA Programme/Generalitat de Catalunya and the Josep Carreras Foundation for institutional support. E.B. was funded by the Spanish Ministry of Science and Innovation (MICINN; grant number PID2020-117212RB-I00/AEI/10.13038/501100011033). E.M.-C. is funded with RESTORE project (EU H2020 Research and Innovation Programme, number 779316) and Spanish projects PI17/01521 and PI20/01313, integrated in the Plan Nacional de I+D+I and co-supported by the ISCIII-Subdirección General de Evaluación and FEDER. O.M.-P. holds an i-PFIS PhD fellowship (grant number IFI17/00034) from Acción Estratégica en Salud 2013–2016 ISCIII, co-financed by Fondo Social Europeo. F.F. holds a PhD fellowship from the INsTRuCT Consortium, which receives. Innovative Training Network subsidy from the EU H2020 program.

AUTHOR CONTRIBUTIONS

F.C.-M. and E.B. conceived and designed the study; F.C.-M., A.G.F.-B., G.G.-T., O.M.-P., L.C., L.B., F.F., and T.L. performed the differentiation,

chromatin immunoprecipitation, co-immunoprecipitation experiments, and immunological assays; F.C.-M. performed the bioinformatic analyses; F.C.-M., A.G.F.-B., G.G.-T., E.M.-C., and E.B. analyzed results; J.R.-U. and E.B. supervised the study; F.C.-M., T.L., and E.B. wrote the manuscript; all authors participated in discussions and interpreting the results.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 29, 2021
Revised: September 27, 2021
Accepted: December 20, 2021
Published: January 18, 2022

REFERENCES

Álvarez-Errico, D., Vento-Tormo, R., Sieweke, M., and Ballestar, E. (2015). Epigenetic control of myeloid cell differentiation, identity and function. Nat. Rev. Immunol. *15*, 7–17.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. Nat. Methods *11*, 1138–1140.

Barragan, M., Good, M., and Kolls, J.K. (2015). Regulation of dendritic cell function by vitamin D. Nutrients *7*, 8127–8151.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E., et al. (2006). High-throughput DNA methylation profiling using universal bead arrays. Genome Res. *16*, 383–393.

Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162.

Bscheider, M., and Butcher, E.C. (2016). Vitamin D immunoregulation through dendritic cells. Immunology *148*, 227–236.

Carlberg, C. (2019). Vitamin D signaling in the context of innate immunity: focus on human monocytes. Front. Immunol. *10*, 2211.

Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics *26*, 2363–2367.

Cauwels, A., and Tavernier, J. (2020). Tolerizing strategies for the treatment of autoimmune diseases: from ex vivo to in vivo strategies. Front. Immunol. *11*, 674.

Chauss, D., Freiwald, T., McGregor, R., Yan, B., Wang, L., Nova-Lamperti, E., Kumar, D., Zhang, Z., Teague, H., West, E.E., et al. (2022). Autocrine vitamin D signaling switches off pro-inflammatory programs of T H 1 cells. Nat. Immunol. *23*, 62–74.

Chu, C.C., Ali, N., Karagiannis, P., Di Meglio, P., Skowera, A., Napolitano, L., Barinaga, G., Grys, K., Sharif-Paghaleh, E., Karagiannis, S.N., et al. (2012). Resident CD141 (BDCA3) + dendritic cells in human skin produce IL-10 and induce regulatory T cells that suppress skin inflammation. J. Exp. Med. *209*, 935–945.

Corzo, C.A., Cotter, M.J., Cheng, P., Cheng, F., Kusmartsev, S., Sotomayor, E., Padhya, T., McCaffrey, T.V., McCaffrey, J.C., and Gabrilovich, D.I. (2009). Mechanism regulating reactive oxygen species in tumor-induced myeloid-derived suppressor cells. J. Immunol. *182*, 5693–5701.

Corzo, C.A., Condamine, T., Lu, L., Cotter, M.J., Youn, J.I., Cheng, P., Cho, H.I., Celis, E., Quiceno, D.G., Padhya, T., et al. (2010). HIF-1α regulates function and differentiation of myeloid-derived suppressor cells in the tumor microenvironment. J. Exp. Med. *207*, 2439–2453.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374.

Dankers, W., Colin, E.M., van Hamburg, J.P., and Lubberts, E. (2017). Vitamin D in autoimmunity: molecular mechanisms and therapeutic potential. Front. Immunol. *7*, 697.

de la Rica, L., Rodríguez-Ubreva, J., García, M., Islam, A.B.M.M.K., Urquiza, J.M., Hernando, H., Christensen, J., Helin, K., Gómez-Vaquero, C., and Ballestar, E. (2013). PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol. *14*, R99.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J. *32*, 645–655.

DeVito, N.C., Plebanek, M.P., Theivanthiran, B., and Hanks, B.A. (2019). Role of tumor-mediated dendritic cell Tolerization in immune evasion. Front. Immunol. *10*, 2876.

Feingold, E.A., Good, P.J., Guyer, M.S., Kamholz, S., Liefer, L., Wetterstrand, K., Collins, F.S., Gingeras, T.R., Kampa, D., Sekinger, E.A., et al. (2004). The ENCODE (ENCyclopedia of DNA elements) project. Science *306*, 636–640.

Ferreira, G.B., Overbergh, L., Verstuyf, A., and Mathieu, C. (2013). 1α,25-Dihydroxyvitamin D3 and its analogs as modulators of human dendritic cells: a comparison dose-titration study. J. Steroid Biochem. Mol. Biol. *136*, 160–165.

Ferreira, G.B., Vanherwegen, A.S., Eelen, G., Gutiérrez, A.C.F., VanLommel, L., Marchal, K., Verlinden, L., Verstuyf, A., Nogueira, T., Georgiadou, M., et al. (2015). Vitamin D3 induces tolerance in human dendritic cells by activation of intracellular metabolic pathways. Cell Rep. *10*, 711–725.

Fetahu, I.S., Höbaus, J., and Kállay, E. (2014). Vitamin D and the epigenome. Front. Physiol. *5*, 164.

Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375.

Garcia-Gomez, A., Li, T., Kerick, M., Català-Moll, F., Comet, N.R., Rodríguez-Ubreva, J., De La Rica, L., Branco, M.R., Martín, J., and Ballestar, E. (2017). TET2- and TDG-mediated changes are required for the acquisition of distinct histone modifications in divergent terminal differentiation of myeloid cells. Nucleic Acids Res. *45*, 10002–10017.

Goudot, C., Coillard, A., Villani, A.C., Gueguen, P., Cros, A., Sarkizova, S., Tang-Huau, T.L., Bohec, M., Baulande, S., Hacohen, N., et al. (2017). Aryl hydrocarbon receptor controls monocyte differentiation into dendritic cells versus macrophages. Immunity *47*, 582–596.e6.

Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G.A., Feber, A., Teschendorff, A.E., Gomez, V., Hergovich, A., Tirabosco, R., Fernanda Amary, M., et al. (2013). Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. Nat. Commun. *4*, 2166.

Haan, S., Keller, J.F., Behrmann, I., Heinrich, P.C., and Haan, C. (2005). Multiple reasons for an inefficient STAT1 response upon IL-6-type cytokine stimulation. Cell. Signal. *17*, 1542–1550.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Ichiyama, K., Chen, T., Wang, X., Yan, X., Kim, B.S., Tanaka, S., Ndiaye-Lobry, D., Deng, Y., Zou, Y., Zheng, P., et al. (2015). The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. Immunity *42*, 613–626.

Jakob, F., Gieseler, F., Tresch, A., Hammer, S., Seufert, J., and Schneider, D. (1992). Kinetics of nuclear translocation and turnover of the vitamin D receptor in human HL60 leukemia cells and peripheral blood lymphocytes-coincident rise of DNA-relaxing activity in nuclear extracts. J. Steroid Biochem. Mol. Biol. *42*, 11–16.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Klug, M., Schmidhofer, S., Gebhard, C., Andreesen, R., and Rehli, M. (2013). 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. Genome Biol. *14*, R46.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program primer3. Bioinformatics *23*, 1289–1291.

Korotkevich, G., Sukhov, V., and Sergushichev, A. (2019). Fast gene set enrichment analysis. BioRxiv. https://doi.org/10.1101/060012.

Kumar, V., Cheng, P., Condamine, T., Mony, S., Languino, L.R., McCaffrey, J.C., Hockstein, N., Guarino, M., Masters, G., Penman, E., et al. (2016). CD45 phosphatase inhibits STAT3 transcription factor activity in myeloid cells and promotes tumor-associated macrophage differentiation. Immunity *44*, 303–315.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lasho, T.L., Tefferi, A., Hood, J.D., Verstovsek, S., Gilliland, D.G., and Pardanani, A. (2008). TG101348, a JAK2-selective antagonist, inhibits primary hematopoietic cells derived from myeloproliferative disorder patients with JAK2V617F, MPLW515K or JAK2 exon 12 mutations as well as mutation negative patients. Leukemia *22*, 1790–1792.

Leek, J.T., Johnson, W.E., Parker, H.S., Fertig, E.J., Jaffe, A.E., Zhang, Y., Storey, J.D., Torres, L.C., 2021. Bioconductor. sva: Surrogate Variable Analysis. R package version 3.42.0.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, T., Garcia-Gomez, A., Morante-Palacios, O., Ciudad, L., Özkaramehmet, S., Van Dijck, E., Rodríguez-Ubreva, J., Vaquero, A., and Ballestar, E. (2020). SIRT1/2 orchestrate acquisition of DNA methylation and loss of histone H3 activating marks to prevent premature activation of inflammatory genes in macrophages. Nucleic Acids Res. *48*, 665–681.

Lio, C.W., Zhang, J., González-Avalos, E., Hogan, P.G., Chang, X., and Rao, A. (2016). Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. Elife *5*, e18290.

Liu, J., Zhang, X., Chen, K., Cheng, Y., Liu, S., Xia, M., Chen, Y., Zhu, H., Li, Z., and Cao, X. (2019). CCR7 chemokine receptor-inducible lnc-Dpf3 restrains dendritic cell migration by inhibiting HIF-1α-Mediated glycolysis. Immunity *50*, 600–615.e15.

Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: in the right place at the right time. Science *361*, 1336–1340.

Mahé, E.A., Madigou, T., Sérandour, A.A., Bizot, M., Avner, S., Chalmel, F., Palierne, G., Métivier, R., and Salbert, G. (2017). Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. Genome Res. *27*, 947–958.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mendes, K., Schmidhofer, S., Minderjahn, J., Glatz, D., Kiesewetter, C., Raithel, J., Wimmer, J., Gebhard, C., and Rehli, M. (2021). The epigenetic pioneer EGR2 initiates DNA demethylation in differentiating monocytes at both stable and transient binding sites. Nat. Commun. *12*, 1556.

Monticelli, S., and Natoli, G. (2017). Transcriptional determination and functional specificity of myeloid cells: making sense of diversity. Nat. Rev. Immunol. *17*, 595–607.

Mora, J.R., Iwata, M., and Von Andrian, U.H. (2008). Vitamin effects on the immune system: vitamins A and D take centre stage. Nat. Rev. Immunol. *8*, 685–698.

Morante-Palacios, O., Fondelli, F., Ballestar, E., and Martínez-Cáceres, E.M. (2021). Tolerogenic dendritic cells in autoimmunity and inflammatory diseases. Trends Immunol. *42*, 59–75.

Müller, F., Scherer, M., Assenov, Y., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2019). RnBeads 2.0: comprehensive analysis of DNA methylation data. Genome Biol. *20*, 55.

Murphy, T.L., Grajales-Reyes, G.E., Wu, X., Tussiwand, R., Briseño, C.G., Iwata, A., Kretzer, N.M., Durai, V., and Murphy, K.M. (2016). Transcriptional control of dendritic cell development. Annu. Rev. Immunol. *34*, 93–119.

Navarro-Barriuso, J., Mansilla, M.J., and Martínez-Cáceres, E.M. (2018). Searching for the transcriptomic signature of immune tolerance induction-biomarkers of safety and functionality for tolerogenic dendritic cells and regulatory macrophages. Front. Immunol. *9*, 2062.

Ong, L.T.C., Schibeci, S.D., Fewings, N.L., Booth, D.R., and Parnell, G.P. (2021). Age-dependent VDR peak DNA methylation as a mechanism for latitude-dependent multiple sclerosis risk. Epigenetics Chromatin *14*, 9.

Pacis, A., Tailleux, L., Morin, A.M., Lambourne, J., MacIsaac, J.L., Yotova, V., Dumaine, A., Danckært, A., Luca, F., Grenier, J.C., et al. (2015). Bacterial infection remodels the DNA methylation landscape of human dendritic cells. Genome Res. *25*, 1801–1811.

Pacis, A., Mailhot-Léonard, F., Tailleux, L., Randolph, H.E., Yotova, V., Dumaine, A., Grenier, J.-C., and Barreiro, L.B. (2019). Gene activation precedes DNA demethylation in response to infection in human dendritic cells. Proc. Natl. Acad. Sci. U S A *116*, 6938–6943.

Pageaud, Y., Plass, C., and Assenov, Y. (2018). Enrichment analysis with EpiAnnotator. Bioinformatics *32*, 1781–1783.

Park, J.H., Van Wyk, H., McMillan, D.C., Quinn, J., Clark, J., Roxburgh, C.S.D., Horgan, P.G., and Edwards, J. (2017). Signal transduction and activator of transcription-3 (STAT3) in patients with colorectal cancer: associations with the phenotypic features of the tumor and host. Clin. Cancer Res. *23*, 1698–1709.

Penna, G., and Adorini, L. (2000). 1α,25-Dihydroxyvitamin D 3 inhibits differentiation, maturation, activation, and survival of dendritic cells leading to impaired alloreactive T cell activation. J. Immunol. *164*, 2405–2411.

Pereira, F., Barbáchano, A., Silva, J., Bonilla, F., Campbell, M.J., Muñoz, A., and Larriba, M.J. (2011). KDM6B/JMJD3 histone demethylase is induced by vitamin D and modulates its effects in colon cancer cells. Hum. Mol. Genet. *20*, 4655–4665.

Piemonti, L., Monti, P., Sironi, M., Fraticelli, P., Leone, B.E., Dal Cin, E., Allavena, P., and Di Carlo, V. (2000). Vitamin D 3 affects differentiation, maturation, and function of human monocyte-derived dendritic cells. J. Immunol. *164*, 4443–4451.

Pratt, J.C., Igras, V.E., Maeda, H., Baksh, S., Gelfand, E.W., Burakoff, S.J., Neel, B.G., and Gu, H. (2000). Cutting edge: gab2 mediates an inhibitory phosphatidylinositol 3'-kinase pathway in T cell antigen receptor signaling. J. Immunol. *165*, 4158–4163.

Ramírez, F., Dündar, F., Sarah Diehl, S., Björn A Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Rasmussen, K.D., Berest, I., Keβler, S., Nishimura, K., Simón-Carrasco, L., Vassiliou, G.S., Pedersen, M.T., Christensen, J., Zaugg, J.B., and Helin, K. (2019). TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. Genome Res. *29*, 564–575.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Sardina, J.L., Collombet, S., Tian, T.V., Gómez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., et al. (2018). Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. Cell Stem Cell *23*, 727–741.e9.

Schaefer, A., Unterberger, C., Frankenberger, M., Lohrum, M., Staples, K.J., Werner, T., Stunnenberg, H., and Ziegler-Heitbrock, L. (2009). Mechanism of Interferon-gamma mediated down-regulation of Interleukin-10 gene expression. Mol. Immunol. *46*, 1351–1359.

Schübeler, D. (2015). Function and information content of DNA methylation. Nature *517*, 321–326.

Segura, E., Touzot, M., Bohineust, A., Cappuccio, A., Chiocchia, G., Hosmalin, A., Dalod, M., Soumelis, V., and Amigorena, S. (2013). Human inflammatory dendritic cells induce Th17 cell differentiation. Immunity *38*, 336–348.

Seuter, S., Neme, A., and Carlberg, C. (2017). Epigenomic PU.1-VDR crosstalk modulates vitamin D signaling. Biochim. Biophys. Acta Gene Regul. Mech. *1860*, 405–415.

Seuter, S., Neme, A., and Carlberg, C. (2018). ETS transcription factor family member GABPA contributes to vitamin D receptor target gene regulation. J. Steroid Biochem. Mol. Biol. *177*, 46–52.

Széles, L., Keresztes, G., Töröcsik, D., Balajthy, Z., Krenács, L., Póliska, S., Steinmeyer, A., Zuegel, U., Pruenster, M., Rot, A., et al. (2009). 1,25-Dihydroxyvitamin D 3 is an autonomous regulator of the transcriptional changes leading to a tolerogenic dendritic cell phenotype. J. Immunol. *182*, 2074–2083.

Teh, B.K., Yeo, J.G., Chern, L.M., and Lu, J. (2011). C1q regulation of dendritic cell development from monocytes with distinct cytokine production and T cell stimulation. Mol. Immunol. *48*, 1128–1138.

Torres-Aguilar, H., Aguilar-Ruiz, S.R., González-Pérez, G., Munguía, R., Bajaña, S., Meraz-Ríos, M.A., and Sánchez-Torres, C. (2010). Tolerogenic dendritic cells generated with different immunosuppressive cytokines induce antigen-specific anergy and regulatory properties in memory $CD4^+$ T cells. J. Immunol. *184*, 1765–1775.

van der Maaten, L. (2014). Accelerating t-SNE using Tree-based algorithms. J. Machine Learn. Res. *15*, 3221–3245.

Van Halteren, A.G.S., Van Etten, E., De Jong, E.C., Bouillon, R., Roep, B.O., and Mathieu, C. (2002). Redirection of human autoreactive T-cells upon interaction with dendritic cells modulated by TX527, an analog of 1,25 dihydroxyvitamin D3. Diabetes *51*, 2119–2125.

Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H., et al. (2015). WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Mol. Cell *57*, 662–673.

Wei, Z., Yoshihara, E., He, N., Hah, N., Fan, W., Pinto, A.F.M., Huddy, T., Wang, Y., Ross, B., Estepa, G., et al. (2018). Vitamin D switches BAF complexes to protect β cells. Cell *173*, 1135–1149.e15.

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu, G.-L., Wang, H., et al. (2016). Cooperative action between SALL4A and TET proteins in stepwise oxidation of 5-methylcytosine. Mol. Cell *64*, 913–925.

Ziegler-Heitbrock, L., Lötzerich, M., Schaefer, A., Werner, T., Frankenberger, M., and Benkhart, E. (2003). IFN-α induces the human IL-10 gene by recruiting both IFN regulatory factor 1 and Stat3. J. Immunol. *171*, 285–290.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Fc Block reagent, human antibody	Miltenyi Biotec	Cat# 130-059-901; RRID: AB_2892112
Anti-human CD14, FITC conjugated (clone TÜK4)	Miltenyi Biotec	Cat# 130-080-701;RRID: AB_244303
Anti-human CD80, PE conjugated (clone 2D10)	Miltenyi Biotec	Cat# 130-097-202; RRID: AB_2659259
Anti-human CD86, APC conjugated (clone FM95)	Miltenyi Biotec	Cat# 130-113-569; RRID: AB_2726174
Anti-human CD11b, APC conjugated (clone ICRF44)	BioLegend	Cat# 301310; RRID: AB_314162
Anti-human CD1a, PE conjugated (clone HI149)	BioLegend	Cat# 300106; RRID: AB_314020
Anti-human HLA-DR, Pe-Cyanine7 conjugated (clone LN3)	Thermo Fisher Scientific	Cat# 25-9956-42; RRID: AB_1582284
LIVE/DEAD™ Fixable Violet Dead Cell Stain Kit	Thermo Fisher Scientific	Cat# L34964
Anti-acetyl-Histone H3 Antibody	Millipore	Cat# 06-599; RRID: AB_2115283
Anti-trimethyl-Histone H3 (Lys27) Antibody	Millipore	Cat# 07-449; RRID: AB_310624
Anti-Trimethyl-Histone H3 (Lys4)	Millipore	Cat# 17-614; RRID: AB_11212770
Rat IgG1 kappa Isotype Control (clone eBRG1)	Thermo Fisher Scientific	Cat# 16-4301-85; RRID: AB_470154
IL-10 Monoclonal Antibody (clone JES3-9D7)	Thermo Fisher Scientific	Cat# 16-7108-85; RRID: AB_469229
IL-6 Monoclonal Antibody (clone MQ2-13A5)	Thermo Fisher Scientific	Cat# 16-7069-85; RRID: AB_469219
Vitamin D receptor	Cell Signaling Technology	Cat# 12550; RRID: AB_2637002
Anti-TET2 antibodody	Abcam	Cat# ab124297; RRID: AB_2722695
Anti-Pstat3 [Y705], (clone 4/P-STAT3)	Fluidigm	Cat# 3158005A; RRID: AB_2811100
Anti-STAT3, (clone 79D7)	Cell Signaling Technology	Cat# 4904; RRID: AB_331269
Anti-pSTAT1 [Y701], (clone 58D6)	Cell Signaling Technology	Cat# 9167; RRID: AB_561284
Anti-STAT1, (clone 42H3)	Cell Signaling Technology	Cat# 9175; RRID: AB_2197984
Anti-pSTAT5 [Y694], (clone 47)	Fluidigm	Cat# 3150005A; RRID: AB_2744690
Anti-STAT5 beta, (clone ST5b-10G1)	Thermo Fisher Scientific	Cat# 13-5300; RRID: AB_2533021
Anti-Lamin B1	Abcam	Cat# ab16048; RRID: AB_443298
Anti-alpha-Tubulin, (clone DM1A)	Sigma-Aldrich	Cat# T6199; RRID: AB_477583
Normal Rabbit IgG Antibody	Millipore	Cat# 12-370; RRID: AB_145841
Normal Mouse IgG Antibody	Millipore	Cat# 12-371; RRID: AB_145840
Biological samples		
Buffy Coats	Catalan Blood and Tissue Bank (CBTB)	Cat# BB014
Chemicals, peptides, and recombinant proteins		
Recombinant Human IL-4	Peprotech	Cat# 200-04; GenPept: P05112
Recombinant Human GM-CSF	Peprotech	Ca# 300-03; GenPept: P04141
1α,25-Dihydroxyvitamin D3	Sigma-Aldrich	Cat# D1530; CAS: 32222-06-3
TG101348, JAK/STAT pathway inhibitor	STEMCELL	Cat# 73472; CAS: 936091-26-8
Potassium perruthenate (VII) (KRuO4)	Thermo Fisher Scientific	Cat# 11877; CAS: 10378-50-4
Triton™ X-100	Sigma-Aldrich	Cat# T8787; CAS: 9036-19-5
Benzonase® Nuclease	Sigma-Aldrich	Cat# E1014; CAS: 9025-65-4
DMP (Dimethyl Pimelimidate)	Thermo Fisher Scientific	Cat# 21666; CAS: 58537-94-3
Critical commercial assays		
MACS CD14 Microbeads	Miltenyi Biotec	Cat# 130-050-201
CD1c (BDCA-1)+ Dendritic Cell Isolation Kit	Miltenyi Biotec	Cat# 130-119-475
Dynabeads™ Untouched™ Human CD8 T Cells Kit	Thermo Fisher Scientific	Cat# 11348D
Cell Trace CFSE	Thermo Fisher Scientific	Cat# C34554

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
APC BrdU Flow Kit	BD Pharmingen	Cat# 552598
ELISA MAXTM Deluxe Set Human IL-10	BioLegend	Cat# 430604
Access IL-6 reagent kit	Beckman-Coulter	Cat# A16369
Maxwell RSC Cultured Cells DNA Kit	Promega	Cat# AS1620
Maxwell RSC simplyRNA cells Kit	Promega	Cat# AS1390
EZ DNA Methylation-Gold Kit	Zymo Research	Cat# D5005
PyroMark Q48 Advanced CpG Reagents	Qiagen	Cat# 974022
Transcriptor First Strand cDNA Synthesis Kit	Roche	Cat# 04897030001
IMMOLASE DNA polymerase Kit	Bioline	Cat# BIO-21047
LightCycler® 480 SYBR Green I Master	Roche	Cat# 0487352001
Infinium MethylationEPIC BeadChip	Illumina	Cat# 20042130
iDeal ChIP-seq kit for Transcription Factors	Diagenode	Cat# C01010055
Magna ChIP™ Protein A+G Magnetic Beads	Sigma-Aldrich	Cat# 16-663
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Sigma-Aldrich	Cat# 11873580001
PureProteome™ Protein G Magnetic Bead System	Sigma-Aldrich	Cat# LSKMAGG02
Micro Bio-Spin® P-6 SSC columns	Bio-Rad	Cat# 7326200
Deposited data		
DNA methylation profile of *in vitro* generated DC and TolDC	This paper	GSE145483
VDR ChIP-Seq	This paper	GSE145584
Oligonucleotides		
Primers for bisulphite pyrosequencing, RT-qPCR and ChIP-qPCR, see Table S2	This paper	N/A
Software and algorithms		
PyroMark Assay Design 2.0 software	Qiagen	Cat# 9019079
Pyromark Q48 Autoprep software	Qiagen	Cat# 9024325
Minfi (R package)	Aryee et al., 2014	https://www.bioconductor.org/packages/release/bioc/html/minfi.html
RnBeads (R package)	Assenov et al. (2014)	https://bioconductor.org/packages/release/bioc/html/RnBeads.html
Limma (R package)	Ritchie et al. (2015)	https://bioconductor.org/packages/release/bioc/html/limma.html
Bowtie2 Aligner v2.2.6	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MarkDuplicates software v1.126	Broad institute	https://broadinstitute.github.io/picard/
Sequence Alignment/Map (SAMtools) v1.2	Li et al. (2009)	https://doi.org/10.1093/bioinformatics/btp352
bamCoverage function (deepTools (v2.0))	Ramírez et al. (2014)	https://deeptools.readthedocs.io/en/2.1.0/content/tools/bamCoverage.html
HOMER Motif Analysis	Heinz et al. (2010)	http://homer.ucsd.edu/homer/motif/
GREAT (version 3.0.0.)	McLean et al. (2010)	http://great.stanford.edu/public/html
EpiAnnotator (R package)	Pageaud et al. (2018)	http://epigenomics.dkfz.de/EpiAnnotator/
fgsea (R package)	Korotkevich et al. (2019)	http://bioconductor.org/packages/release/bioc/html/fgsea.html
tSNE	van der Maaten, 2014	https://github.com/lvdmaaten/bhtsne/
sva (R package)	(Leek et al., 2021)	https://bioconductor.org/packages/release/bioc/html/sva.html
DoRothEA (R package)	Garcia-Alonso et al. (2019)	https://bioconductor.org/packages/release/data/experiment/html/dorothea.html
Nichenetr (R package)	Browaeys et al. (2020)	https://github.com/saeyslab/nichenetr

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Esteban Ballestar (eballestar@carrerasresearch.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- DNA methylation and ChIP-seq data for this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible through GEO Series accession numbers GSE145483 and GSE145584.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Differentiation of TolDCs and DCs from peripheral blood monocytes

For *in vitro* differentiation experiments, we obtained buffy coats from anonymous donors through the Catalan Blood and Tissue Bank (CBTB). The CBTB follows the principles of the World Medical Association (WMA) Declaration of Helsinki. The Committee for Human Subjects of Bellvitge Hospital approved the study (PR275/17). Given the anonymous nature of the volunteers, no information about the gender and age was provided by the CBTB. Before providing the first blood sample, all donors received detailed oral and written information, and signed a consent form at the CBTB. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Paque gradient centrifugation. MOs were isolated from PBMCs using positive selection with MACS CD14 microbeads (Miltenyi Biotec). Cells were resuspended in RPMI Medium 1640 + GlutaMAXTM-1 (Gibco, Life Technologies) containing 10% fetal bovine serum, 100 units/mL penicillin, and 100 μg/mL streptomycin. For TolDC differentiation, the medium was supplemented with 10 ng/mL human IL-4, 10 ng/mL GM-CSF (PeproTech), and 10 nM of vitamin D3 or calcitriol (Sigma Aldrich). For DCs, the medium did not contain vitamin D. Anti-IL-6 1 μg/mL (Invitrogen) and rat IgG isotype 1 μg/mL (eBioscience) was added during differentiation process when required. In some cases, specified in the text, vitamin D3 was added for 30 min following differentiation to DCs (DC + vitD). In other experiments, differentiation was performed in the presence of a JAK2 inhibitor (TG101348, STEMCELL) at 500 nM.

Isolation and culture of peripheral blood DCs

For the validation of our differentiation model, peripheral blood CD1c + DCs were isolated and cultured for three days with or without 100 nM vitamin D (Sigma Aldrich). These samples were also obtained from anonymous donors through the Catalan Blood and Tissue Bank (CBTB). The CBTB follows the principles of the World Medical Association (WMA) Declaration of Helsinki. The Committee for Human Subjects of Bellvitge Hospital approved the study (PR275/17). Given the anonymous nature of the volunteers, no information about the gender and age was provided by the CBTB. Before providing the first blood sample, all donors received detailed oral and written information, and signed a consent form at the CBTB. For this, PBMCs were isolated by Ficoll-Paque gradient centrifugation followed by a CD1c + DCs purification with CD1c (BDCA-1)+. Dendritic Cell Isolation Kit (Miltenyi Biotec) according to manufacturer instructions. Obtained cells were cultured at a concentration of 10^6 cells/mL in RPMI Medium 1640 + GlutaMAXTM-1 (Gibco, Life Technologies) containing 10% fetal bovine serum and 100 units/mL penicillin with or without vitamin D.

METHOD DETAILS

$CD8^+$ cell proliferation assay

Allogenic $CD8^+$ T-cells isolated using negative selection with the human CD8 T Cells Kit (Invitrogen) were labeled with carboxyfluorescein succinimidyl ester (CFSE) and seeded in 96-well plates at 200,000 cells/well, with TolDCs or DCs at different ratios (TolDC/DC:CD8+ T-cell ratios: 1:2, 1:4, and 1:6). $CD8^+$ cells were then stimulated with anti-CD3/CD28 Dynabeads 5 μL/mL (Invitrogen) and cultured for 5 days. Anti-IL-6 1 μg/mL (Invitrogen), anti-IL-10 1 μg/mL (eBioscience) and rat IgG isotype 1 μg/mL (eBioscience) was added during co-culture process when required. $CD8^+$ T-cell proliferation was analyzed by FACS and determined by considering the proliferating $CD8^+$ T-cells those where CFSE staining had decreased compared to unstimulated $CD8^+$ T-cells.

BrdU proliferation assay

MOs were differentiated to DCs and TolDCs as described above and BrdU (APC BrdU Flow kit, BD Pharmingen) pulses were added to a final concentration of 10 μM at days 2 and 4. On days 3, 4 and 5 cells were harvested and 10^6 cells were prepared for flow cytometry as described by the manufacturer. In brief, cells were fixed for 30 minutes on ice, permeabilized for 5 minutes on ice and

treated with DNAse for 1 h at 37°C to expose incorporated BrdU. Cells were then stained with fluorescent anti-BrdU antibody for 20 minutes at room temperature and analyzed in a BD FACSCanto-II flow cytometer. The HAFTL pre-B cell line was used as control for proliferation.

Flow cytometry

For the study of surface cell markers, cells were harvested after differentiation culture and washed once with PBS. Cell staining was performed in a staining buffer (PBS with 4% fetal bovine serum and 0.4% EDTA) after blocking for non-specific binding with Fc block (BD Pharmingen) for 5 minutes on ice. Cells were stained for 20 minutes on ice. Antibodies used included: CD14-FITC, CD80-PE, CD86-APC (Miltenyi biotec), CD11b-APC, CD1a-PE (Biolegend), HLA-DR-PeCy7 (eBioscience). Cells were also stained with the viability dye LIVE/DEAD™ Fixable Violet (Invitrogen) according to manuacturer's conditions. After staining, cells were fixed with PBS + 4% paraformaldehyde and analyzed in a BD FACSCanto-II flow cytometer in the following 48 h.

Cytokine measurements

For *in vitro* experiments, the concentration of IL-10 cytokine was measured from the cell culture supernatants using an enzyme-linked immunosorbent assay (ELISA), according to the manufacturer's instructions (BioLegend, San Diego, CA, USA). The concentration of IL-6 was measured with Beckman DXI Immunoassay analyzer using the access IL-6 reagent kit (Beckman Coulter) following the manufacturer's instructions.

Genomic DNA and total RNA extraction

DNA was extracted with a Maxwell RSC Cultured Cells DNA kit (Promega) following manufacturer's instructions. Similarly, total RNA was extracted with Maxwell RSC simplyRNA cells kit (Promega) following manufacturer's instructions.

Bisulfite (BS) and oxidative-bisulfite (oxBS) pyrosequencing

500 ng of genomic DNA was BS-converted with an EZ DNA Methylation-Gold kit (Zymo Research), following the manufacturer's instructions. The oxBS samples were purified via buffer exchange with Micro Bio-Spin® P-6 SSC columns (Bio-Rad Laboratories, Inc., USA) and eluted in ~22 μL MilliQ-water. After DNA denaturation with 1.25 μL NaOH (1M) for 30 min in a shaking incubator at 37°C, DNA was oxidized with 2 μL KRuO4 (15 mM) (Alfa Aesar, Germany) for 60 min in an ice-water bath (vortexing the reaction twice) and centrifuged at 16000 g for 15 min. Finally, oxidized DNA were BS converted using the EZ DNA Methylation™ kit (Zymo Research, CA, USA). BS- and oxBS-treated DNA was PCR-amplified using IMMOLASE DNA polymerase kit (bioline). Primers were designed with PyroMark Assay Design 2.0 software (Qiagen) (see Table S3 for primer sequences). Finally, PCR amplicons were pyrosequenced with the PyroMark Q24 system and analyzed with PyroMark Q48 Autoprep (Qiagen). 5mC levels were derived from the oxBS data, while 5hmC levels were calculated by subtracting the oxBS values from the BS values (5mC+5hmC) using the same biological replicate, as described in (Garcia-Gomez et al., 2017).

Real-time quantitative Polymerase Chain Reaction (RT-qPCR)

250 ng of total RNA were converted to cDNA with Transcriptor First Strand cDNA Synthesis Kit (Roche) following manufacturer's instructions. RT-qPCR primers were designed with Primer3 software (Koressaar and Remm, 2007) (see Table S3). RT-qPCR reactions were prepared with LightCycler® 480 SYBR Green I Master (Roche) according to manufacturer's instructions and analyzed with a LightCycler® 480 instrument (Roche).

Western blot

Protein expression and downregulation was visualized by western blotting, performed using standard Western blot. The following antibodies were used for Western blotting and Co-immunoprecipitation: Anti-Vitamin D3 Receptor (Cell Signaling), anti-TET2 (Abcam), anti-pStat3 (Fluidigm), anti-Stat3 (Cell Signaling Technology), anti-pStat1 (Cell Signaling Technology), anti-Stat1 (Cell Signaling Technology), anti-pStat5 (Fluidigm), anti-Stat5 (Thermo Fisher Scientific), anti-Lamin B1 (Abcam), anti-aTubulin (Sigma-Aldrich). anti-rabbit IgGs (Merck Millipore) and anti-mouse IgGs (Merck Millipore).

Co-immunoprecipitation (Co-IP)

Co-IP assays were performed using TolDCs differentiated from CD14 + monocytes for 3 days. Cell extracts were prepared in lysis buffer [50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl, 1% Triton-X-100, protease inhibitor cocktail (cOmplete™, Merck)] with corresponding units of Benzonase (Sigma) and incubated at 4°C for 4 h. 100 μL of supernatant was saved as input and diluted with 2× Laemmli sample buffer (5x SDS, 20% glycerol, 1M Tris–HCl (pH 8.1)). Supernatant was first incubated with PureProteome™ Protein A/G agarose suspension (Merck Millipore) for 1 h to remove background signal. The lysate was then incubated overnight at 4°C with respective crosslinked primary antibody. The cross-linking was performed in 20 mM dimethyl pimelimidate (DMP) (Pierce, Thermo Fisher Scientific, MA, USA) dissolved in 0.2 M sodium borate (pH 9.0). Subsequently, the beats were quenched with 0,2M of ethanolamine (pH 8.0) and resuspended at 4°C in PBS until use. Beads were then washed three times with lysis buffer at 4°C. Sample elution was done by acidification using a buffer containing 0.2 M glycine (pH 2.3) and diluted with 2× Laemmli. Samples and inputs were denatured at 95°C in the presence of 1% β-mercaptoethanol.

DNA methylation profiling

Infinium MethylationEPIC BeadChip (Illumina, Inc., San Diego, CA, USA) arrays were used to analyze DNA methylation. This platform allows >850,000 methylation sites per sample to be interrogated at single-nucleotide resolution, covering 99% of the reference sequence (RefSeq) genes. The samples were bisulfite-converted using EZ DNA Methylation-Gold™ Kit (Zymo Research, Irvine, CA, USA) and were hybridized in the array following the manufacturer's instructions. Image processing and intensity data extraction software and procedures were as previously described (Bibikova et al., 2006). Each methylation data point was obtained from a combination of the Cy3 and Cy5 fluorescent intensities from the methylated and unmethylated alleles. Background intensity computed from a set of negative controls was subtracted from each data point. For representation and further analysis, we used beta and M values. The beta value is the ratio of the methylated probe intensity to the overall intensity (the sum of the methylated and unmethylated probe intensities). It can take a value between 0 and 1, and was used to derive heatmaps and to compare DNA methylation percentages from bisulfite-pyrosequencing experiments. The M value is calculated as the $\log_2$ ratio of the intensities of the methylated versus unmethylated probes. For the purpose of statistical analysis, M values are more suitable because they are normally distributed.

Raw methylation data were preprocessed with the minfi package (Aryee et al., 2014). Data quality was assessed using the minfi and RnBeads packages (Aryee et al., 2014; Assenov et al., 2014; Müller et al., 2019). After Snoob normalization, data were analyzed using aneBayes moderate t test available in the limma package (Ritchie et al., 2015). Several criteria have been proposed as representing significant differences in methylated CpGs, but in this study we considered a probe to be differentially methylated if it had a methylation differential of 20% and if it was significant ($q < 0.05$).

ChIP-seq analysis

Chromatin immunoprecipitation was performed using the iDeal ChIP-seq kit for Transcription Factors (Diagenode), according to the manufacturer's instructions. Briefly, cells on day 3 of differentiation were cross-linked with 1% formaldehyde for 15 min and glycine was added to quench the reaction (final concentration 125 mM, incubated for 5 min at room temperature). Cells were washed once with cold PBS, scraped off the plates, and pelleted. To obtain a soluble chromatin extract, cells were resuspended in 1 mL LB1 (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and 1× complete protease inhibitor) and incubated while rotating at 4°C for 10 min. Samples were centrifuged, resuspended in 1 mL LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 1× complete protease inhibitor) and incubated while rotating at 4°C for 10 min. Finally, samples were centrifuged, resuspended in 1 mL LB3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% *N*-lauroylsarcosine, 1% Triton X-100 and 1× complete protease inhibitor). Chromatin extracts were sonicated for 12.5 min using a Covaris M220 focused ultrasonicator at a peak power of 75, and a duty factor of 10 and 200 cycles per burst. The lysates were incubated with anti-VDR antibody (12,550, Cell Signaling) bound to 30 μL protein A or protein G Dynabeads and incubated overnight at 4°C, keeping 5% as input DNA. Magnetic beads were sequentially washed with low-salt buffer (150 mM NaCl, 0.1% SDS, 1% Triton X-100, 1 mM EDTA and 50 mM Tris-HCl), high-salt buffer (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 1 mM EDTA and 50 mM Tris-HCl), LiCl buffer (150 mM LiCl, 0.5% sodium deoxycholate, 0.1% SDS, 1% Nonidet P-40, 1 mM EDTA and 50 mM Tris-HCl) and TE buffer (1 mM EDTA and 10 mM Tris-HCl). For ChIP-seq, beads were resuspended in elution buffer (1% SDS, 50 mM Tris-HCl pH 8.0, 10 mM EDTA and 200 mM NaCl) and incubated for 30 min at 65°C. After centrifugation, the eluate was reverse-cross-linked overnight at 65°C. The eluate was then treated with RNaseA for 1 h at 37°C and with Proteinase K (Roche) for 1 h at 55°C and the DNA was recovered using a Qiagen PCR purification kit.

Sequencing reads from ChIP-seq experiments were mapped to the hg19 assembly of human reference genome using Bowtie2 Aligner v2.2.6 (Langmead and Salzberg, 2012). After removing reads with MAPQ < 30 with Sequence Alignment/Map (SAMtools) v1.2 (Li et al., 2009), PCR duplicates were eliminated using the Picard function available in MarkDuplicates software v1.126. Peak calling was determined using SPP (with parameters –npeak=300000 –savr –savp -rf). The irreproducible discovery rate (IDR) was used to filter peaks (IDR < 0.05). To visualize individual ChIP-seq data on Integrative Genomics Viewer (IGV), we converted bam output files to normalized bigwig format using the bamCoverage function in deepTools (v2.0).

ChIP-qPCR

ChIP assays were performed as previously described (Li et al., 2020). Briefly, MOs, DCs and TolDCs were crosslinked with 1% methanol-free formaldehyde (Thermo Fisher) for 15 min and subjected to immunoprecipitation after sonication. ChIP experiments were performed using the LowCell# ChIP kit™ protein A (Diagenode, Liège, Belgium). We used antibodies against vitamin D3 Receptor (Cell Signaling), acetylated H3 (H3ac), trimethylated lysine 27 of histone H3 (H3K4me3Millipore) and trimethylated lysine 4 of histone H3. Corresponding rabbit IgG (Diagenode) is used as control. Protein binding was analyzed by real-time quantitative PCR, and data are represented as ratio of the enriched fraction with respect to input. ChIP primers were designed for the areas flanking differentially methylated CpGs and their sequences are shown in Table S3.

Microarray reanalysis

Affymetrix datasets from human monocytes, and from *in vitro*- and *in-vivo* DCs and MACs were obtained from GSE40484 (Segura et al., 2013) and GSE102046 (Goudot et al., 2017). Affymetrix raw data from MOs, DCs and TolDCs were obtained from GSE13762 (Széles et al., 2009). Affymetrix raw data files were normalized by the robust multiarray average (RMA) algorithm and summarized,

after background correction, using the R package oligo (Carvalho and Irizarry, 2010). Normalized expression datasets were then merged and corrected for batch effects using ComBat function of the sva package. Finally, t-Distributed Stochastic Neighbor Embedding (tSNE) of 1000 most variant genes was performed in R using Rtsne package.

Data analysis

Hierarchical clustering was carried out based on Pearson correlation distances and average linkage criteria. For low-dimensional analysis, we used principal component analysis (PCA). Transcription-factor motifs were enriched for each set using HOMER software v4.10.3. Specifically, we used the findMotifsGenome.pl algorithm (with parameters -size 200 -cpg) to search for significant enrichment against a background sequence adjusted to have similar CpG and GC contents. Genomic regions for genetic context location were annotated using the annotatePeaks.pl algorithm in the HOMER v4.10.3 software application (Heinz et al., 2010). To determine the location relative to a CpG island (CGI), we used 'hg19_cpgs' annotation in the annotatr v1.8 R package. GREAT software (McLean et al., 2010) was used to enrich downstream pathways and gene ontologies. We used the single nearest gene option to identify associations between genomic regions and genes. Chromatin state analysis for DCs were assessed using the EpiAnnotator R package (Pageaud et al., 2018). Inference of TF activities from expression values were calculated using DoRothEA (Garcia-Alonso et al., 2019). We used the nichenetr package (Browaeys et al., 2020) to predict ligand activity.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were done in R v3.5.1. Data distributions were tested for normality. Normally distributed data were tested using two-tailed unpaired Student's t-tests; non-normal data were analyzed with the appropriate non-parametric statistical test. Levels of significance are indicated as: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. Non-significance ($P \geq 0.05$) is indicated as 'ns'.

Published online 5 May 2021 *Nucleic Acids Research, 2021, Vol. 49, No. 9* ***5057–5073***
doi: 10.1093/nar/gkab322

Activation-induced deaminase is critical for the establishment of DNA methylation patterns prior to the germinal center reaction

Francesc Català-Moll[1,2], Anna G. Ferreté-Bonastre[1,2,†], Tianlu Li[1,2,†], Dieter Weichenhan[3], Pavlo Lutsik[3], Laura Ciudad[1,2], Ángel F. Álvarez-Prado[4], Javier Rodríguez-Ubreva[1,2], Christian Klemann[5], Carsten Speckmann[5,6], Amaya Vilas-Zornoza[7], Hassan Abolhassani[8], Mónica Martínez-Gallo[9], Romina Dieli-Crimi[9], Jacques G. Rivière[10,11,12], Andrea Martín-Nalda[10,11,12], Roger Colobran[9,12,13], Pere Soler-Palacín[10,11,12], Sven Kracker[14], Lennart Hammarström[8], Felipe Prosper[7], Anne Durandy[14], Bodo Grimbacher[5,15,16,17,18], Christoph Plass[3] and Esteban Ballestar[1,2,*]

[1]Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain, [2]Chromatin and Disease Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Barcelona, Spain, [3]Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, [4]B Cell Biology Laboratory, Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), 28029 Madrid, Spain, [5]Institute for Immunodeficiency, Center for Chronic Immunodeficiency (CCI), Medical Center, Faculty of Medicine, University of Freiburg, 79110 Freiburg, Germany, [6]Faculty of Medicine, Center for Pediatrics and Adolescent Medicine, Medical Center, University of Freiburg, Germany, [7]Center for Applied Medical Research (CIMA), University of Navarra, 31008 Pamplona, Spain, [8]Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE 14186 Stockholm , Sweden, [9]Immunology Division, Hospital Universitari Vall d'Hebron and Diagnostic Immunology Research Group, Vall d'Hebron Research Institute (VHIR), Barcelona, Spain, [10]Pediatric Infectious Diseases & Immunodeficiencies Unit, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute (VHIR), Autonomous University of Barcelona, Barcelona, Spain, [11]Infection in Immunocompromised Pediatric Patients Research Group, Vall d'Hebron Institut de Recerca (VHIR), Hospital Universitari Vall d'Hebron, 08035 Barcelona, Spain, [12]Jeffrey Modell Diagnostic and Research Center for Primary Immunodeficiencies, Barcelona, Spain, [13]Department of Cell Biology, Physiology and Immunology, Universitat Autònoma de Barcelona (UAB), Bellaterra, Catalonia, Spain, [14]Laboratory of Human Lymphohematopoiesis, Imagine Institute, INSERM UMR 1163, Université de Paris, Paris, France, [15]German Center for Infection Research (DZIF), Satellite Center Freiburg, Germany, [16]Centre for Integrative Biological Signalling Studies (CIBSS), Albert-Ludwigs University, Freiburg, Germany, [17]RESIST, Cluster of Excellence 2155 to Hanover Medical School, Satellite Center Freiburg, Germany and [18]Institute of Immunity & Transplantation, Royal Free Hospital, University College London, UK

Received February 07, 2020; Revised April 15, 2021; Editorial Decision April 16, 2021; Accepted April 19, 2021

ABSTRACT

Activation-induced deaminase (AID) initiates antibody diversification in germinal center B cells by deaminating cytosines, leading to somatic hypermutation and class-switch recombination. Loss-of-function mutations in AID lead to hyper-IgM syndrome type 2 (HIGM2), a rare human primary antibody deficiency. AID-mediated deamination has been proposed as leading to active demethylation of 5-methycytosines in the DNA, although evidence both supports and casts doubt on such a role. In this study, using whole-genome bisulfite sequencing of HIGM2 B cells, we investigated direct AID involvement in active DNA demethylation. HIGM2 naïve and memory B cells both display widespread DNA methy-

*To whom correspondence should be addressed. Tel: +34 935572800; Fax: +34 934651472; Email: eballestar@carrerasresearch.org
†The authors wish it to be known that, in their opinion, the second and third authors should be regarded as Joint Second Authors.
Present address: Christian Klemann, Department of Paediatric Pulmonology, Allergy and Neonatology, Hannover Medical School, Hannover, Germany.

lation alterations, of which ∼25% are attributable to active DNA demethylation. For genes that undergo active demethylation that is impaired in HIGM2 individuals, our analysis indicates that AID is not directly involved. We demonstrate that the widespread alterations in the DNA methylation and expression profiles of HIGM2 naïve B cells result from premature overstimulation of the B-cell receptor prior to the germinal center reaction. Our data support a role for AID in B cell central tolerance in preventing the expansion of autoreactive cell clones, affecting the correct establishment of DNA methylation patterns.

INTRODUCTION

Hyper-IgM syndrome type 2 (HIGM2) is a rare primary antibody deficiency, with autosomal recessive inheritance, characterized by loss-of-function mutations in activation-induced deaminase (AID) (1), an enzyme required for several crucial steps of B cell terminal differentiation. AID converts deoxycytosines (dCs) into deoxyuracils (dUs), producing dU:dG mismatches that are removed by mismatch repair and base-excision repair (2). Deaminase activity is required for somatic hypermutation (SHM) and class-switch recombination (CSR) of immunoglobulin (Ig) genes, which are necessary processes for affinity maturation and antibody diversification within the germinal centers (GC) (3,4). AID deficiency results in the absence of CSR and SHM, and leads to lymphoid hyperplasia (1). HIGM2 patients have normal or elevated serum IgM levels with severe reduction of IgG, IgA and IgE, resulting in considerable susceptibility to bacterial infections (1).

In addition to its role in CSR and SHM, AID has been proposed to participate in active DNA demethylation through deamination of 5-methylcytosine (5mC), leading to a mismatch that is converted to G:C by thymine DNA glycosylase (TDG), followed by base-excision repair. The potential role of AID in active DNA demethylation was first proposed by Petersen-Mahrt and colleagues (5). Initial studies in this topic mainly focusing on non-lymphoid cells, such as including zebrafish embryos and heterokaryon-based reprogramming, supported that model (6–8). These studies subsequently led to its study in B cells, given that activated B cells display the highest levels of AID expression (9). During the past decade, conflicting reports have both supported and discounted a role in active demethylation for AID in that context [reviewed in (10)]. For instance, Fritz *et al.* (11) performed reduced-representation bisulfite sequencing (RRBS) of mouse splenic naïve B cells from wild type and AID-deficient mice, activated *ex vivo* for 72 h and found no significant differences in their DNA methylation profiles. Similar conclusions were obtained looking at GC B cells from AID-deficient mice and using MethylCap-Seq (12). However, more recently, Dominguez and colleagues showed that the transit of B cells through the GC is associated with marked locus-specific loss of methylation and increased methylation diversity, both of which are lost in *Aicda*−/− animals (13). Methodological aspects could explain the discrepancies between these studies, including the limited coverage and resolution of the three aforementioned analyses, or the study of different cells (*in vitro* activated or isolated GC B cells).

On the other hand, different *in vitro* studies suggest that 5mC is a poorer substrate than C, although early studies showed that human AID can deaminate 5mC (5). For instance, Abdouni and colleagues showed that the efficient deamination of 5mC by zebrafish AID is due to the flexibility of its structure, in comparison with that of other AID orthologs, including human AID (14). In addition, the comparison of AID with other members of the AID/APOBEC family have shown that human AID deaminates 5mC only weakly because the 5-methyl group fits poorly in its DNA-binding pocket (15). Larijani *et al.* have shown that methylated-CpG motifs, but not their unmethylated counterparts, are in fact protected from AID-mediated deamination (16). Nabel and colleagues have also shown that steric requirements for cytosine deamination are one intrinsic barrier to the proposed function of deaminases in DNA demethylation (17).

In parallel, the discovery of alternative enzymatic pathways that lead to *bona fide* active DNA demethylation through ten-eleven translocation methylcytosine dioxygenase (TET)-mediated oxidation of methylcytosines (18,19) raised more doubts about the possibility that AID redundantly plays such a role. There is currently no consensus about whether AID is involved in mediating DNA demethylation in specific cell contexts.

Whole-genome analysis has shown the occurrence of a vast amount of demethylation associated with B cell differentiation. Changes occur mostly during naïve B cell activation, yielding memory B cells (20,21) that coincide with the highest peak of AID expression (3). Naïve B cells start to proliferate upon activation by antigen encounter. Then they express AID which triggers the secondary diversification of antibodies by SHM and CSR. This is followed by affinity maturation which finally leads to (a) a new cycle of SHM or (b) terminal differentiation into memory or plasma B cells depending on the affinity of the B cell receptor (BCR) for the cognate antigen (22). It has been originally reported that AID is targeted to Ig genes with the involvement of RNA polymerase II (23). Nevertheless, it has since been described that AID is also recruited to non-Ig genes to mediate recurrent mutations and these genes are collectively termed 'AID off-targets' (24,25).

In this study, we took advantage of the exceptional possibility to investigate the direct role of AID in active demethylation by comparing the complete DNA methylomes of naïve and memory B cells of HIGM2 patients with those of healthy individuals. By studying two sibling patients with a homozygous mutation for AID that results in a severely truncated enzyme, we were able to determine its direct link with DNA methylation defects and infer its catalytic activity in relation to active DNA demethylation.

Our results show that the absence of AID catalytic activity affects DNA methylation in naïve and memory B cells. The majority of the changes occurring in the transition from naïve to memory B cells arise from passive demethylation and are linked to late-replicating domains. However, for those potentially associated with active demethylation, we found no evidence of direct involvement of AID, and our analysis suggests that TET enzymes are responsi-

ble for DNA methylation changes in this cell context. The increased DNA demethylation noted in naïve B cells of HIGM2 patients is associated with premature demethylation of BCR downstream genes prior to the GC reaction. Indeed, we found that these changes are related to the expansion of autoreactive clones, which suggests a major role for AID in preventing the expansion of such clones under normal conditions.

MATERIALS AND METHODS

Human samples

The patients included in the study fulfilled the diagnostic criteria for hyper-IgM syndrome type 2, based on ESID clinical diagnostic criteria (26) and genetic confirmation of *AICDA* mutation and exclusion of other primary and secondary causes of immunodeficiencies. Samples come from the Medical Center of the University Hospital, University of Freiburg, Freiburg, Germany and Hospital Universitari Vall d'Hebron, Barcelona, Spain. The Committees for Human Subjects of the local hospitals approved the study, which was conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. All samples were in compliance with the guidelines approved by the local ethics committee and all donors (and/or their parents) received oral and written information about the possibility that their blood would be used for research purposes.

Isolation of B cell populations. Peripheral blood mononuclear cells (PBMCs) were obtained from blood. After Ficoll-Isopaque density centrifugation (Rafer, Zaragoza, Spain), collected cells were washed twice with ice-cold PBS, followed by centrifugation at 2000 rpm for 5 min. Next, cells were labeled with antibodies to CD19 – FITC (Miltenyi Biotec, clone LT19), CD27 – APC (Miltenyi Biotec, clone M-T271), IgD – PE (SouthernBiotech, Cat. No. 2032-09) and IgM – PerCP/Cy5.5 (BioLegend, clone MHM-88) for 20 min on ice in a staining buffer (PBS with 4% FBS and 2 mM EDTA). Naïve B cells ($CD19^+$ $CD27^-$ IgD^+) and unswitched memory B cells ($CD19^+$ $CD27^+$ IgD^+) were obtained by FACS sorting on a MoFlo Astrios (Beckman Coulter). Purified samples were pelleted and stored at −80°C.

For isolation of naïve autoreactive B cells. Total B cells were isolated from PBMCs using positive selection with MACS CD19 microbeads (Miltenyi Biotec). Next, cells were stained with CD27-APC (Miltenyi Biotec, clone M-T271), IgD – PE (SouthernBiotech, Cat. No. 2032-09), HLA-DR – PE-Cy7 (eBioscience, clone LN3), 9g4 primary ab (igm Bioscience) and donkey anti-rat IgG (H + L) – Alexa Fluor 488 (invitrogen). 9g4+ naïve B cells ($CD27^-$ IgD^+ $9g4^+$) and 9g4- naïve B cells ($CD27^-$ IgD^+ $9g4^-$) were obtained by FACS sorting on a BD FACSAria II (BD Biosciences). Purified samples were pelleted and stored at −80°C.

Genomic DNA extraction. For whole-genome bisulfite sequencing, DNA was extracted with a QIAamp DNA micro kit (Qiagen) according to the manufacturer's protocol. For pyrosequencing analysis (detailed below), DNA was extracted with a Maxwell RSC Cultured Cells DNA kit (Promega), following the manufacturer's protocol.

Tagmentation-based whole-genome bisulfite sequencing. For whole-genome bisulfite sequencing, 30 ng of genomic DNA was used to produce four independent barcoded sequencing libraries per DNA sample using the tagmentation method (27). Sequencing of the TWGBS libraries was done on a HiSeq 2000, PE 125 bp mode. Bisulfite sequencing reads were processed by the DKFZ bisulfite analysis workflow. In brief, the reads were trimmed using Trimmomatic, pre-processed and aligned using MethylCTools, with default parameters (V. Hovestadt, S. Picelli, B. Radlwimmer, M.Z. and P.L., unpublished data), which uses the Burrows-Wheeler alignment algorithm (28). Following quality control of bisulfite conversion (>99.5% in all samples) and of read-mapping (80–90% could be mapped on average), we performed methylation calling using methylCtools. A summary of the sequencing data for each sample is provided in Supplementary Table S1.

Differentially methylated region (DMR) calling. DMRs were detected with the DeNovoDMR algorithm included in the Specific Methylation Analysis and Report Tool (SMART2) (29) using all the default parameters except for the number of CpGs per segment, which was set to 4, the absolute mean methylation difference, which was set to 0.2, and a threshold value of P of 0.01. Only those CpGs with a coverage of ≥5 in all samples were considered in the construct of the SMART input matrix. DMR calling was performed on all possible comparisons between naïve and memory B cells for both control and HIGM2 patients.

Bisulfite pyrosequencing. 500 ng of genomic DNA was converted with an EZ DNA Methylation-Gold kit (Zymo Research), following the manufacturer's instructions. Bisulfite-treated DNA was PCR amplified using primers (see Supplementary Table S2) designed with PyroMark Assay Design 2.0 software (Qiagen). Finally, PCR amplicons were pyrosequenced with the PyroMark Q24 system and analyzed with PyroMark CpG software (Qiagen).

Processing of chromatin immunoprecipitation (ChIP) assays with sequencing (ChIP-Seq) datasets

Sequencing reads from ChIP-seq experiments from the BLUEPRINT consortium (21) were mapped to the hg19 assembly of human reference genome using Burrows-Wheeler Aligner (BWA) v0.7.13 (with parameters -q 5, -l 32, -k 2). After removing reads with MAPQ <30 with Sequence Alignment/Map (SAMtools) v1.2, PCR duplicates were eliminated using the Picard function available in MarkDuplicates software v1.126. Peak calling was performed using macs2 (with parameters -p 1e-2 –nomodel –shift 0 -B –SPMR). Only peaks with an overlap of ≥0.5 between replicates were considered. Histone mark signals around DMR sets were extracted with the annotatePeaks.pl algorithm available in Hypergeometric Optimization of Motif EnRichment (HOMER) software v4.10.3 (with parameters: size = 10 000, hist = 10).

Super-enhancer identification. H3K27ac ChIP-seq data from BLUEPRINT database were used to identify the super-enhancer regions, as described previously (30) using Rank-Ordering of Super-Enhancers (ROSE) software. An enhancer stitching distance of 15 kb was used along with a 2.5 kb transcriptional start site (TSS)-exclusion window.

DNA methylation data analysis. Hierarchical clustering was carried out based on Pearson correlation distance metrics and average linkage criteria. For low-dimensional analysis, we used the *t*-distributed stochastic neighbor embedding (*t*-SNE) method implemented in the Rtsne v0.15 package.

Transcription factor motifs were enriched for each set of DMRs using HOMER software v4.10.3. Specifically, we used findMotifsGenome.pl algorithm (with parameters -size given -cpg) to search for significant enrichment against a background sequence adjusted to have similar CpG and GC contents.

Transcription factor binding analysis was performed interrogating the overlap between the different sets of DMRs with ChIP-seq data for transcription factors available for GM12878 lymphoblastoid cell line from the ENCODE Project (31). The enrichment factor was calculated against random regions as a background, and *P* values were calculated using Fisher's exact test. Finally, the transcription factors downstream of the BCR signaling pathway were manually annotated from a curated database (32).

Chromatin states and histone mark enrichments analysis for NBC, GC B cells and ncsMBC were assessed using a custom adaptation of the EpiAnnotator R package (33) using BLUEPRINT data (21). DMRs were converted to hg38 assembly with the liftOver function in the rtracklayer v1.42 R package.

Replication timing data in the GM12878 lymphoblastoid cell line were obtained from the UW Repli-seq track of the UCSC Genome Browser. Genomic replication timing values were binned in deciles where the first decile contained the regions of latest replication and the last decile the regions of earliest replication. Identified DMRs were then overlapped to these regions by genomic location.

DMR annotation for genetic context location was performed using the annotatePeaks.pl algorithm in the HOMER software v4.10.3. To determine the location relative to a CpG island (CGI), we used 'hg19_cpgs' annotation in the annotatr v1.8 R package.

GREAT software (34) was used to enrich downstream pathways and gene ontologies. We used the single nearest gene option for the association between genomic regions with genes.

Genes obtained with HOMER annotation software were used to study correlations between their associated DMRs and gene expression data from healthy naïve, memory and GC B cells from BLUEPRINT database, as well as with the AID off-target genes dataset obtained from Álvarez-Prado *et al*. (35).

All statistical analysis (excluding T-WGBS and ChIP-seq analyses) were done in R v3.5.1. Data distributions were tested for normality. Normal data were tested using two-tailed unpaired Student's t-tests; non-normal data were analyzed with the appropriate non-parametric statistical test. Levels of significance are indicated as: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$. Non-significance ($P \geq 0.05$) was indicated as 'ns'.

Public RRBS of B cell activation. Data of EBV and CD40/IL4 B cell activation were downloaded from the NCBI Gene Expression Omnibus (GSE49629) (36). Methylation calls from RRBS data were filtered so that only those CpGs with a minimum of five reads per position in all samples were retained. Since RRBS genomic coverage is significantly lower than T-WGBS we only tested the methylation status of positions common to two datasets.

BCR activation and EBV infection. For naïve B cell activation and infection experiments, we obtained buffy coats from anonymous donors through the Catalan Blood and Tissue Bank (CBTB). The CBTB follows the principles of the World Medical Association (WMA) Declaration of Helsinki. Before providing the first blood sample, all donors received detailed oral and written information and signed a consent form at the CBTB. PBMCs were isolated using Ficoll-Paque gradient centrifugation. Total B cells were isolated from PBMCs using positive selection with MACS CD19 microbeads (Miltenyi Biotec). Next, cells were stained with CD27-APC (Miltenyi Biotec, clone M-T271) and IgD – PE (SouthernBiotech, Cat. No. 2032–09) and naïve B cells were sorted as $CD27^{-}IgD^{+}$. For EBV infection, pure naïve B cells were incubated with B95–8 cell supernatant for 3 h at 37ºC in order to infect them with EBV, they were harvested after 30 days. For BCR activation, pure naïve B cells were cultured in RPMI medium 1640 GlutaMAXTM-1 (Gibco, Life Technologies) containing 20% fetal bovine serum (Gibco, Life Technologies), 1% sodium pyruvate and 1% Penicillin Streptomycin (Gibco Thermo Fisher Scientific, MA, USA). Cells were activated with 10 μg/ml anti-IgM (SouthernBiotech), 0.1 μg/ml MEGACD40L® Protein (Enzo) and 50 ng/ml IL21 (Tebubio) for 24 h (ChIP) or 7 days (bisulfite pyrosequencing).

RNA-sequencing. RNA-sequencing libraries from total mRNA were prepared and purified using Illumina Stranded mRNA Prep, Ligation kit (Illumina). Sequencing was performed on a NextSeq500 (Illumina) with a NextSeq 500/550 High Output Kit v2.5 (75 Cycles, Illumina, 20024906) to generate 38-nucleotide paired-end reads at a read depth of at least 20 million reads per sample.

RNA-seq data analysis. For bioinformatics analyses, raw sequence reads were initially processed using FastQC (Babraham Institute, Cambridge, UK) for quality control, and then adapter sequences and poor quality reads were removed using TrimGalore that uses Cutadapt for trimming. Quality-filtered reads were then mapped to the human genome (hg19) using STAR, and only the uniquely mapped reads were kept. Read counts were calculated using HTSeq-count. Differentially expressed genes were identified using R package DESeq2 (fold change ≥ 2 and adjusted *P*-value < 0.05). Inference of TF activities from expression values were calculated using DoRothEA (37). Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to perform functional enrichment analysis of differentially expressed genes.

ChIP-PCR. Cell-sorter isolated naïve B cells from healthy donors were activated as described above for 24 h. 1×10^6 unstimulated and stimulated B cells were crosslinked with 1% methanol-free formaldehyde (Thermo Fisher) for 10 min subjected to 15 min of sonication utilizing Covaris M220 Focused-ultrasonicator (Covaris, MA, USA). ChIP experiments were performed using the LowCell# ChIP kit™ protein A, according to manufacturer's instructions (Diagenode, Liege, Belgium). Anti-BATF antibody (#8638) and corresponding rabbit IgG (#2729) were both obtained from Cell Signaling. Protein binding was analyzed by real-time quantitative PCR, and data are represented as a ratio of the enriched fraction with respect to input. ChIP primers are shown in Supplementary Table S2.

RESULTS

Study strategy

We obtained peripheral blood from two sibling HIGM2 patients, both with the same homozygous mutation for the *AICDA* gene, and two sex-matched healthy controls. Specifically, the patients carried a deletion (Exon 2 c.22_40del19) that generates a frameshift variant (p.Arg8Asnfs*19) that affects the DNA binding region of AID located within its nuclear localization signal domain (Figure 1A). This deletion impairs both CSR and SHM (38). We confirmed the impact on CSR by inspecting the peripheral B cell compartment by flow cytometry. The two HIGM2 patients were characterized by the absence of class-switched memory B cells (csMBC; $CD19^+CD27^+IgM^-IgD^-$), as described (1) (Figure 1B). In contrast, classic non-class switched memory B cells (ncsMBC; $CD19^+CD27^+IgM^+IgD^+$) and naïve B cells (NBC; $CD19^+CD27^-$IgM+IgD+) were present in patients (Figure 1B). It has been shown that, under physiological conditions, ncsMBC cells display certain levels of SHM at the immunoglobulin locus (20), which supports the expression of AID during their maturation in GCs. Hence, the comparison between DNA methylation profiles of NBC and ncsMBC isolated from healthy and HIGM2 individuals is an adequate model to investigate the potential role of AID in DNA demethylation.

HIGM2 patients display an aberrant methylation profile in naïve and unswitched memory B cells

We performed tagmentation-based whole-genome bisulfite sequencing (T-WGBS), a version of the WGBS method that allows analysis of limited DNA amounts (27), in NBC and ncsMBC (henceforth referred to as 'naïve' and 'memory' B cells) isolated from two HIGM2 siblings and two sex-matched healthy controls (Figure 1C). Pearson correlation and *t*-distribution stochastic neighbor embedding (t-SNE) between samples were highly reproducible between replicates with correlation coefficient to be >0.9 (Supplementary Figure S1A, B). We also compared our DNA methylation data from healthy controls with public data from the International Cancer Genome Consortium (ICGC) (39) and Oakes *et al.* (20), in which we observed high correlation (Pearson correlation >0.85), thereby confirming the robustness of our data (Supplementary Figure S1C, D).

Global inspection of DNA methylation confirmed previous studies (20,21) that the transition from naïve to memory B cells is accompanied by global demethylation of the genome in healthy controls (Figure 1D). However, HIGM2 patients showed a partial impairment of demethylation during naïve-to-memory B cell differentiation (Figure 1D, E), which is compatible with a potential role of AID as a demethylating enzyme. The transition from naïve to memory B cells in healthy controls is characterized by the demethylation of 30175 DMRs. In contrast, comparing naïve to memory B cells in HIGM2 patients only identified 4803 hypomethylated DMRs (Supplementary Table S3). Furthermore, we also observed that naïve B cells were more demethylated in HIGM2 patients than in healthy controls (a total of 2936 hypomethylated DMRs) (Figure 1F). Taken together, these global observations suggest that AID loss not only affects the DNA methylation patterns in the transition from naïve to memory B cells but also has a potential role in establishing the B cell DNA methylome in earlier stages of development.

A high proportion of DNA demethylation events identified in HIGM2 are due to passive demethylation of late-replicating domains

Recent studies have shown that a high proportion of the demethylation events occurring in cancer and in differentiation processes are associated with high proliferation rates (40). Such demethylation events take place in regions known as 'partially methylated domains' (PMDs), which are characterized by repressive chromatin and low-GC density (40), and do not appear to occur in regions of high methylation (HMD: highly methylated domains). PMDs encompass regions that undergo late replication and their demethylation is a passive event, which is a result of inefficient DNA remethylation during DNA replication (36,40,41). Recent reanalysis of the B cell lineage DNA methylation profiles published by the BLUEPRINT consortium (21) has shown the occurrence of demethylation of PMDs in the transition towards memory B cells and antibody-secreting plasma cells (42). This highlights the importance of separating DNA methylation analysis into PMD and non-PMD regions when interrogating the occurrence of active demethylation processes in order to exclude those changes due to DNA replication-dependent or passive demethylation.

To address this matter, we first identified total differentially methylated regions (DMRs) by performing pairwise comparisons of DNA methylation between all samples. Specifically, we compared naïve to memory B cells in both control and HIGM2 patients (control_NBC versus control_ncsMBC, HIGM2_NBC versus HIGM2_ncsMBC), as well as comparing HIGM2 patients to controls for both naïve and memory B cells (control_NBC versus HIGM2_NBC, control_ncsMBC versus HIGM2_ncsMBC). We then divided the DMRs into PMD and HMD regions, which were previously identified by Zhou *et al.* (40) to be present either in all cell types (common) or cell-type specific (partial). We found that the majority of DMRs overlapped with PMDs that were either common or partially shared by cell types (72.5%, Figure 2A). We then classified DMRs into two groups: the first com-

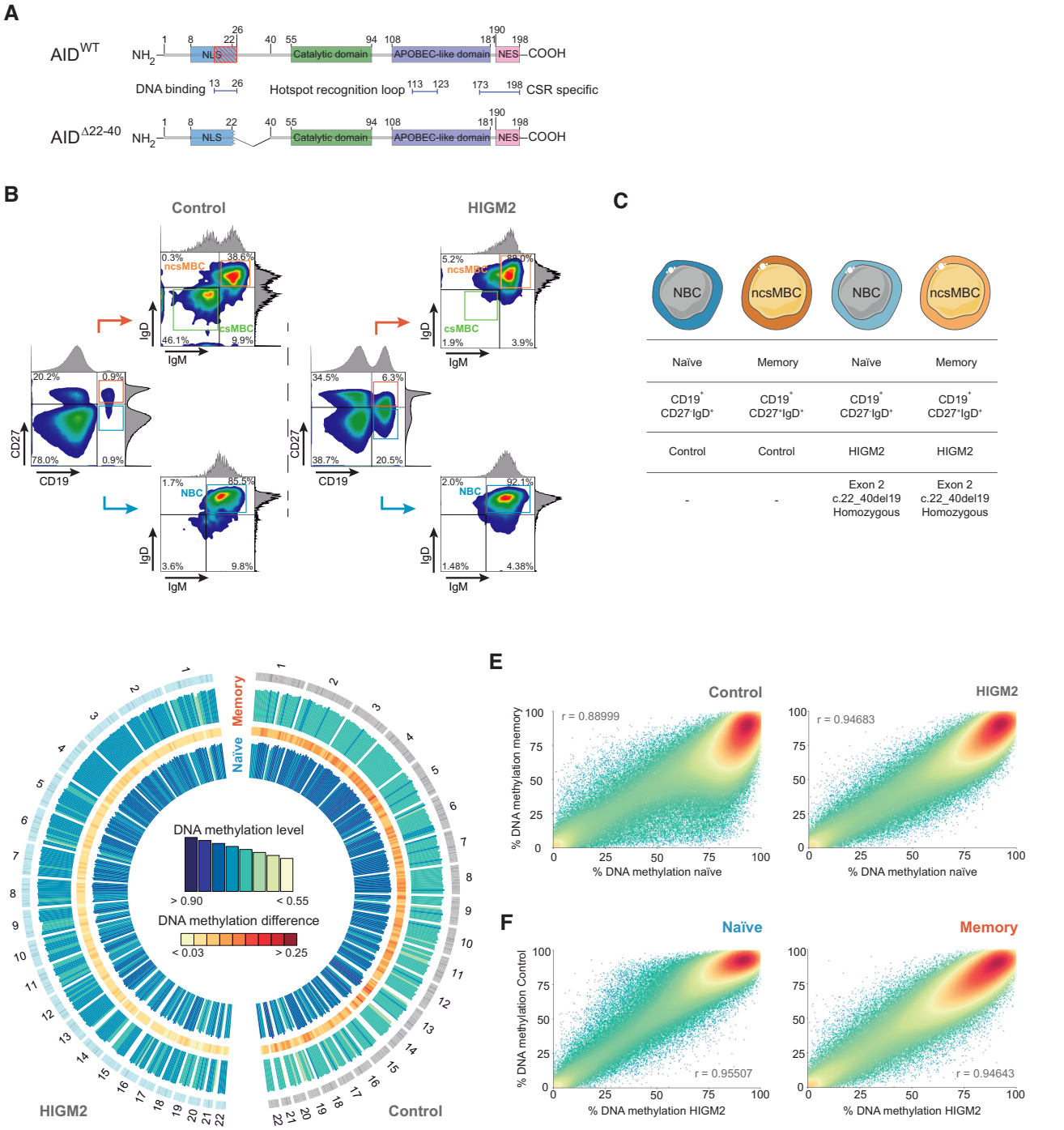


Figure 1. DNA methylomes of HIGM2 B cell subpopulations determined by T-WGBS. (**A**) Graphical representation of wild type (WT) *AICDA* gene and truncated AID of HIGM2 patients used in this study. The mutation in HIGM2 AID harbors a deletion from 22–40 which encompasses the DNA binding domain. (**B**) Representative examples of cell sorting strategy of B cell populations (naïve B cells, NBC; classic non-class-switched memory B cells, ncsMBC; class-switched memory B cells, csMBC). (**C**) Description of the B cell subpopulations analyzed, including surface markers and details of mutation. (**D**) Circos plot of DNA methylation of naïve B cells (inner circle) and unswitched memory B cells (outer circle) for controls (right) and HIGM2 patients (left). Histogram tracks represent the average methylation levels of 10 Mb windows. Heatmap shows the DNA methylation differences between naïve and memory B cells (**E**) Density scatter plots of average DNA methylation in windows of 500 bp in naïve and memory B cells in controls (left) and HIGM2 patients (right). (**F**) Density scatter plots of average DNA methylation in windows of 500 bp comparing HIGM2 and control for naïve (left) and memory (right) B cells.

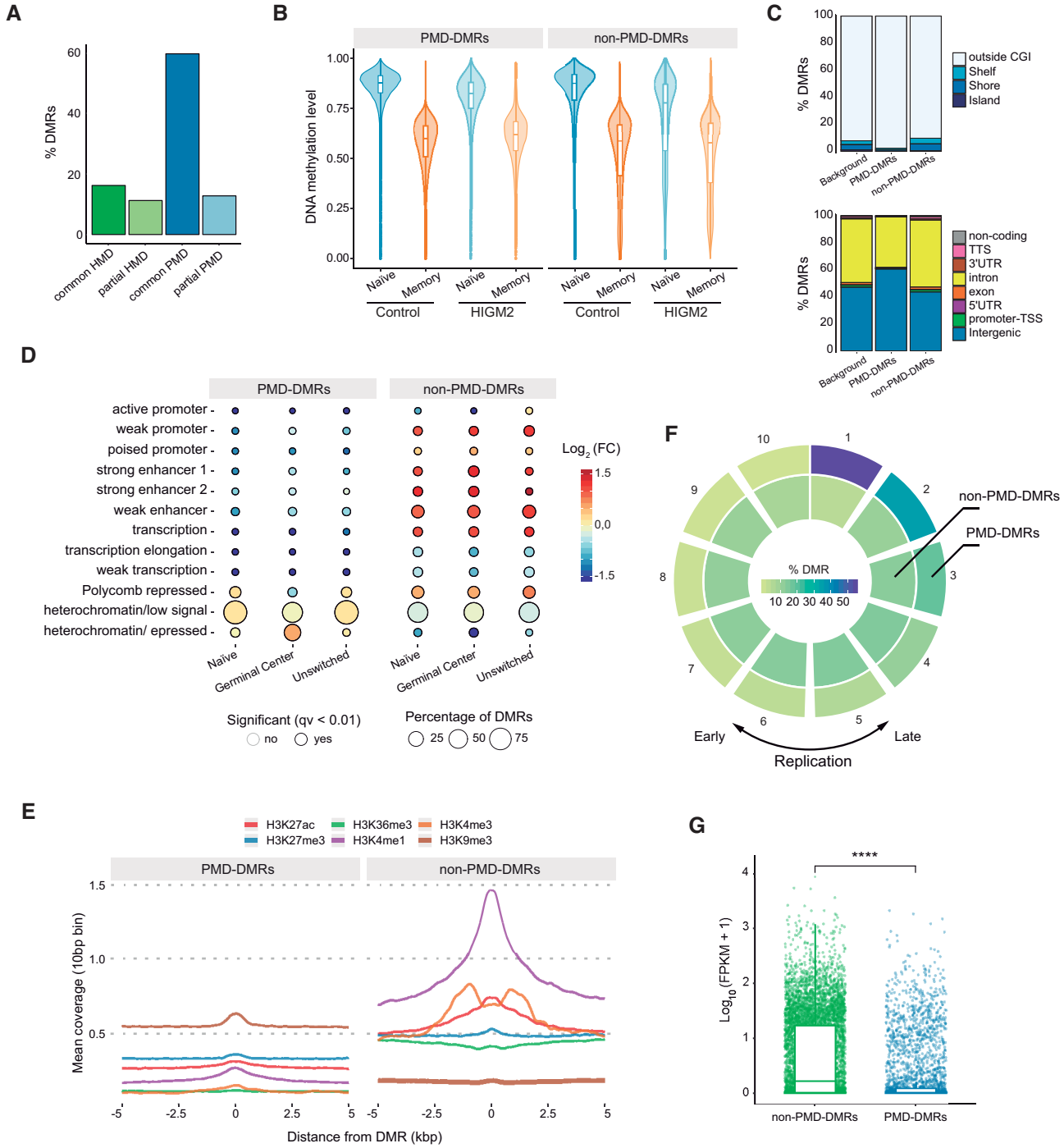


Figure 2. Detection and characterization of partially methylated regions. (**A**) Bar plot showing the percentage of DMRs in PMD and HMD (partially and highly methylated domains, respectively) regions annotated by Zhou and colleagues (40). 'Common' refers to PMD and HMD regions common to all cell types analyzed by Zhou and colleagues. 'Partial' refers to those regions detected only in a fraction of the cell types analyzed. (**B**) Box and violin plots summarizing the distribution of DNA methylation levels for controls and HIGM2 patients, in which the DMRs were characterized as inside (PMD-DMRs) and outside (non-PMD-DMRs) of PMDs. (**C**) Location proportions of PMD-DMRs and non-PMD-DMRs in the context of CpG islands (CGIs) and gene-related regions. (**D**) Bubble chart depicting the enrichment (red) or depletion (blue) of chromatin states defined from ChromHMM ChIP-seq data of healthy naïve, GC and unswitched memory B cells. Color scale represents the logarithmic fold change, circle size indicates the percentage of DMRs in the chromatin state, and the edge color indicates the statistical significance of the enrichment (black: significant, none: not significant; $q < 0.01$). (**E**) ChIP-seq data of the represented histone marks in healthy GC B cells were downloaded from the BLUEPRINT database and mean coverage was calculated for 10bp bins ±5 kb centered around the midpoints of PMD-DMRs (left) and non-PMD-DMRs (right). (**F**) Circular representation of the proportion of PMD- (outer circle) and non-PMD-DMRs (inner circle) in deciles of replication timing data of GM12878 cell line. Color scale represents the proportion of DMRs in each decile. The first and last deciles correspond to the regions of latest and earliest replication, respectively. (**G**) Box and dot plots showing the distribution of expression values of associated genes of PMD- and non-PMD-DMRs. FPKM refers to fragments per kb of transcript per million mapped reads. Expression data from healthy GC B cells downloaded from BLUEPRINT database. Statistical tests: two-tailed Fisher's exact (D) and unpaired Wilcoxon's (G) tests (**** *P*-value < 0.0001).

prising DMRs that coincided with common PMDs (PMD-DMRs), and the second containing DMRs that did not overlap with common PMDs (non-PMD-DMRs). To validate the identity of these PMD-DMRs, inspection of the methylation values showed that the two groups of DMRs had intermediate methylation values in memory B cells, although non-PMD-DMRs had slightly lower methylation levels (Figure 2B). Utilizing functional genomic feature analysis, we found that most PMD-DMRs were located outside of CpG islands and in intergenic regions (Figure 2C). In relation to chromatin states, we observed that DMRs occurring in PMDs were mainly associated with heterochromatic regions, while non-PMD-DMRs were highly enriched at enhancers and active promoters, in concordance to what was previously described as generally associated with replication-independent demethylation (43) (Figure 2D). This was also confirmed by overlapping DMRs with ChIP-seq data of healthy GC B cells obtained from the BLUEPRINT database, in which non-PMD-DMRs were enriched in enhancer marks, including H3K4me1, H3K4me3 and H3K27ac (Figure 2E). It has been previously described that PMDs are characterized by their association with late-replication domains (40), hence we analyzed DMRs based on their association with late or early phase replication. First, genomic density data obtained from GM12878 Repli-seq (UW Repli-seq, UCSC Genome Browser) were divided into deciles by their replication timing and then overlapped by genomic location with identified DMRs. We confirmed that PMD-DMRs were mainly found in late-replicating regions (Figure 2F) and were accompanied by lower expression of associated genes in GC B cells (Figure 2G). Taking all these observations into account, our results suggest that most of the DNA methylation changes occur in *bona fide* PMDs. However, the existence of a set of non-PMD-DMRs (∼27%) located in highly active regions suggests the potential participation of DNA replication-independent or, in other words, active demethylation events, and AID could directly drive these events.

HIGM2-associated defects in DNA methylation in the transition from naïve to memory B cells do not have the features of AID targets

Given all the previous considerations, including the removal of DNA methylation changes related to DNA replication (PMD-DMRs), our model allows us to examine whether AID has a direct role in mediating demethylation in the B cell lineage. In this context, GC B cells, in the transition from naïve to memory, displayed the highest *AICDA* mRNA levels (Supplementary Figure S2A) (as obtained from the BLUEPRINT database). Hence, it is plausible that defective AID in HIGM2 patients could play a role in aberrant DNA demethylation occurring during the transition from naïve to memory B cells within the GC.

To explore this possibility, we first identified DMRs whose DNA demethylation could potentially be driven by AID (P-AID DMRs). These DMRs are demethylated in the transition from naïve to memory B cells in healthy controls but not in HIGM2 patients. Furthermore, DMRs that were already aberrantly demethylated in naïve HIGM2 B cells compared to controls were excluded (Figure 3A). A total of 522 DMRs (containing 450 different genes) fulfilled these conditions (Supplementary Table S4).

We then compared P-AID DMRs associated genes with data obtained from a mouse model that allows the identification of targets of AID deamination events (35), in which 271 non-Ig genes were identified to be dependent on AID catalytic activity and referred to as AID off-target genes. We found low correspondence between these two sets of genes in which only 6 genes (corresponding to 6 DMRs) overlap with the identified AID off-target genes (Figure 3B). We also tested for the presence of described AID hotspot sequences (RGYW and WRCY) (44) in our DMRs and found a significant increase for the WRCY hotspot with respect to the background, although it appeared too low (<1 hotspot per 100 bp) to be of biological relevance (Figure 3C). On the other hand, we found that DMRs associated with AID off-target genes underwent demethylation in HIGM2 patients (Supplementary Figure S2B).

Two recent studies have characterized the genomic and epigenomic features of AID off-target regions in mice. Independently, they found that AID targets regions with convergent transcription from intragenic super-enhancers (45,46). In this sense, there were no significant differences regarding gene localization between the DMRs associated with AID off-targets and P-AID DMRs (Supplementary Figure S2C and Supplementary Table S4). However, AID off-target DMRs exhibited greater enrichment of enhancer regions, as observed by both ChromHMM (Figure 3D) and histone mark (Supplementary Figure S2D) analyses, and were associated with more transcriptional activity of associated genes in GC B cells (Figure 3E). We also found that, although the two DMR groups had a similar percentage of overlap with super-enhancers (P-AID 21%, AID off-target 33.3%; Figure 3F), the super-enhancers of the AID off-targets had a stronger signal for H3K27ac (Figure 3G, H; Supplementary Figure S2E) and greater transcriptional activity of their associated genes than P-AID (Supplementary Figure S2F). Finally, we hypothesized that if AID had a role mediating active DNA demethylation, we would expect to see differences in CpG sites containing WRCY hotspots at the super-enhancers of AID off-target genes. However, such differences were not observed (Supplementary Figure S2G).

Taken together, our results suggest that, despite the differences in DNA methylation associated with B cell activation between wild type and AID-deficient B cells, if we assume that mechanisms and targets are conserved between mouse and human, such demethylation may not be directly associated with AID catalytic activity.

While the proposed mechanism of DNA demethylation by AID implies that the 5mC conversion in thymine could be repaired and replaced with an unmethylated cytosine (5), the removal of methyl groups of cytosines mediated by TET enzymes involves the generation of oxidation intermediates, including 5hmC (18) (Supplementary Figure S3A). It is therefore plausible that AID-dependent demethylation would associate with lower levels of 5hmC compared to TET-dependent demethylation. Hence, we analyzed publicly-available DNA methylation data obtained from GC and naïve B cells of *Aicda*$^{-/-}$ mice (13). We selected three sets of CpGs: (i) CpGs demethylated in the transition

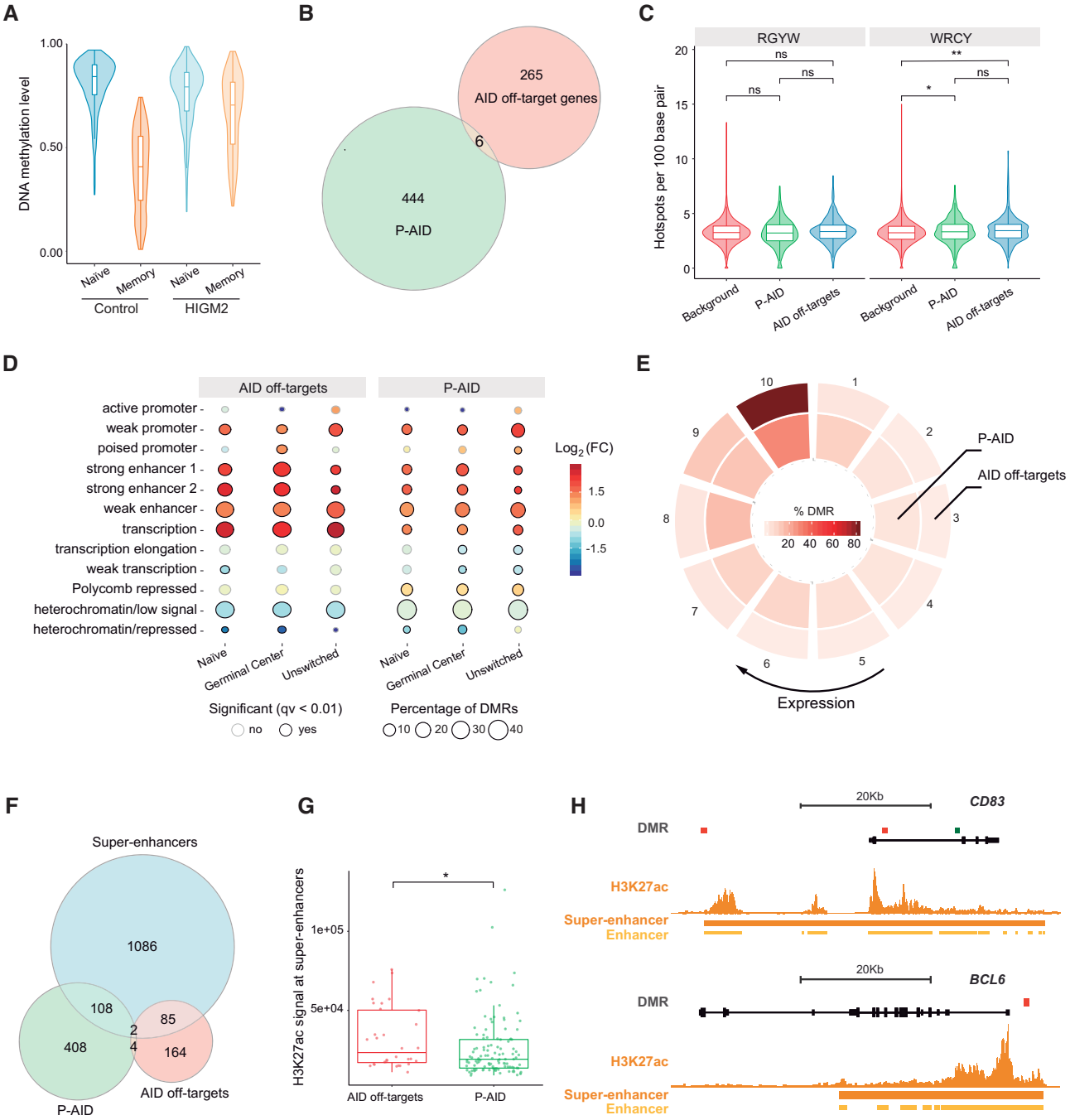


Figure 3. Indirect involvement of AID in DNA demethylation dynamics during B cell activation. (**A**) Box and violin plot representation of DNA methylation of DMRs that are potentially demethylated directly by AID (P-AID). These DMRs were filtered by the presence of demethylation between control naïve and memory B cells which were not present in HIGM2 B cells. Furthermore, DMRs that displayed demethylation in HIGM2 naïve B cells compared to controls were excluded. (**B**) Venn diagram showing the overlap between associated genes of P-AID DMRs and the human orthologues of mouse AID off-target genes as defined by Álvarez-Prado *et al.* (35). (**C**) Box and violin plots displaying the frequency of WRCY/RGYW (W = A/T; R = G/A; Y = C/T) hotspots per 100 bp in DMRs of each subset. Background was generated as 100bp windows of the whole genome. (**D**) Bubble plot depicting the enrichment (red) or depletion (blue) of chromatin states. Color represents the logarithmic-fold change, size shows the percentage of DMRs in the chromatin state, and the edge indicates the statistical significance of the enrichment (black: significant, none: not significant, q-value < 0.01). Chromatin state data was obtained from ChromHMM, generated in GC B cells. (**E**) Circular representation of proportion of DMRs in deciles of genes whose expression is ranked from low to high expression in GC B cells. The first and last deciles correspond to the least and most highly expressed genes, respectively. Color scale represents the proportion of DMRs in each decile. Gene expression data was obtained from the BLUEPRINT database. (**F**) Venn diagram of the overlap between P-AID and AID off-target DMRs with super-enhancers in GC B cells. Super-enhancer identification is described in the Materials section utilizing H3K27ac ChIP-seq data from the BLUEPRINT database. (**G**) Box plots showing H3K27ac ChIP-seq peaks of super-enhancers associated with AID off-target (red) and P-AID (green) DMRs. One-sided unpaired Wilcoxon's test was used to examine signal intensity differences. (**H**) DNA methylation and H3K27ac profiles in the vicinity of two representative genes. DMR color indicates the type of DMR, either P-AID (green) or AID off-target DMRs (red). H3K27ac ChIP-seq signals from GC B cells are shown in dark orange. Super-enhancers and enhancers are depicted. Statistical tests: two-tailed unpaired Wilcoxon's (C) and Fisher's exact (D) tests (* P-value < 0.05, ** P-value < 0.01; ns is not significant).

from naïve to GC B cells in WT but not in *Aicda*$^{-/-}$ mice, termed mouse potential AID targets (mouse P-AID); (ii) CpGs significantly demethylated in naïve-to-GC transition in both WT and knockout mice and, therefore, more likely to be TET-dependent, termed positive control; (iii) CpGs that did not present demethylation in the transition from naïve to GC B cells neither in WT nor in *Aicda*$^{-/-}$ mice, termed negative control (differential DNA methylation < 0.05) (Supplementary Figure S3B). Next, we merged this data with public hydroxy-meDIP-seq data of mouse B cell activation (47) to determine the hydroxymethylation status of these three groups of CpGs. We found that mouse P-AID CpGs presented similar levels of hydroxymethylation to those that are TET-dependent (positive control) compared to negative control (Supplementary Figure S3C). On the other hand, we only found 21 genes associated with mouse P-AID CpGs that overlapped with the aforementioned AID off-target genes (18), which represents <8% of the AID-off target genes (Supplementary Figure S3D). Taken together, these results suggest that, at least in a mouse model, the majority of demethylation associated with B cell differentiation may not be directly driven by AID activity.

AID deficiency results in premature demethylation of the BCR pathway of naïve B cells

Our initial analysis suggested that alterations in DNA methylation not only occurred in the transition from naïve to memory B cells, but were already present in naïve B cells of HIGM2 individuals in comparison with healthy controls. Specifically, HIGM2 naïve B cells appeared to be more demethylated than those of healthy controls (Figures 1F and 4A). AID expression has customarily been associated with the GC reaction (3), however, more recent evidence suggests that AID may have a role in earlier stages of B cell development (48,49).

Comparing naïve B cells from HIGM2 to control, we detected 2152 hypomethylated DMRs (Figure 4A) and 127 hypermethylated DMRs (Supplementary Figure S4A), excluding PMDs. Both hypo- and hypermethylated DMRs were mostly found outside CpG islands in intergenic regions and introns (Supplementary Figure S4B). Genetic annotation of DMRs in CpG islands reveals an association with promoters (Supplementary Figure S4C and Supplementary Table S5). However, while hypomethylation was associated with enhancer regions, hypermethylation was mainly enriched in promoters (Supplementary Figure S4D), in agreement with its reported regulatory role in the 'spurious' initiation of transcription (50).

We observed that the DMRs that were hypomethylated in naïve B cells from HIGM2 patients underwent normal demethylation during the transition from naïve to memory cells in healthy controls (Figure 4A). Our results suggest that the demethylation of these DMRs may be essential for correct naïve-to-memory B cell differentiation, and premature demethylation observed in HIGM2 naïve B cells may indicate aberrant pre-activation outside of the GC. This is consistent with the finding that genes associated with these DMRs became upregulated during the activation of B cells in the GC (Figure 4B) and were associated with functional categories related to B cell activation via BCR (Figure 4C). Some genes that have altered DMRs, such as *BATF* (Figure 4D) and *MEF2A* (Supplementary Figure S4E), are crucial to B cell development (47,51).

To explore the possibility that aberrant DNA methylation in HIGM2 naïve B cells may be due to pre-activation outside the GC, we first analyzed the enrichment for TF binding motifs with HOMER software in HIGM2 naïve DMRs. Some of the most enriched TFs are downstream of the BCR pathway, including BATF, IRF8 and RELA (32) (Supplementary Figure S4F). We then validated these results through enrichment analysis of the ChIP-seq data available for GM12878 cells from the ENCODE consortium (31) and we again observed enrichment of TFs downstream of BCR activation in DMRs found to be altered in naïve B cells of HIGM2 patients (Figure 4E).

The type III latency state of the Epstein-Barr virus (EBV) is characterized by the constitutive activation of the BCR and CD40 pathways (52,53), both of which are major signaling pathways essential for B cell activation (54). In that sense, the B lymphoblastoid cell line GM12878 presents a type III latency state (55) and is therefore a good model to investigate the involvement of BCR/CD40 in aberrant DNA methylation of naïve B cells in HIGM2 patients. Firstly, we represented DNA methylation levels of publicly-available data of EBV-transformed and CD40L/IL-4-activated B cells (36) in regions that overlapped with DMRs identified in naïve B cells of HIGM2 patients and observed that EBV transformation effectively reproduced the aberrant DNA demethylation presented in HIGM2 patients (Figure 4F). Conversely, such changes did not take place when activation was performed with CD40/IL-4 alone, suggesting that the BCR pathway has a significant role in driving aberrant DNA methylation (Figure 4F). Furthermore, these DNA methylation effects were not reproduced in hypermethylated DMRs (Supplementary Figure S5A). To confirm these results, we transformed naïve B cells isolated from healthy donors with EBV and, after 30 days, we interrogated DNA methylation of the set of CpGs that we had previously identified to be aberrantly demethylated in HIGM2. We observed that these CpGs underwent demethylation following EBV-mediated transformation of naïve B cells (Supplementary Figure S5B). Additionally, we directly stimulated BCR in isolated healthy naïve B cells and observed similar demethylation of the same CpGs, which supported the direct role of BCR in their demethylation (Figure 4G). Taken together, our findings suggest that alterations in DNA methylation of HIGM2 naïve B cells may at least be partly due to the aberrant pre-activation of the BCR at some point of B cell development prior to the GC reaction.

Since ChIP-seq analysis in GM12878 cells identified BATF to bind to almost 50% of regions that correspond to DMRs in HIGM2 naïve B cells (Figure 4E), we therefore hypothesize that BATF might be the main mediator of demethylation in naïve-to-memory transition, and its aberrant recruitment may drive the aberrant demethylation observed in HIGM2 naïve B cells. BATF, together with IRF4 (56,57), is a regulator of B and T cell activation via the BCR and TCR pathways respectively. We observed that 87% of the ChIP-seq peaks of IRF4 overlapped with BATF peaks in GM12878 (Figure 4H). As ex-

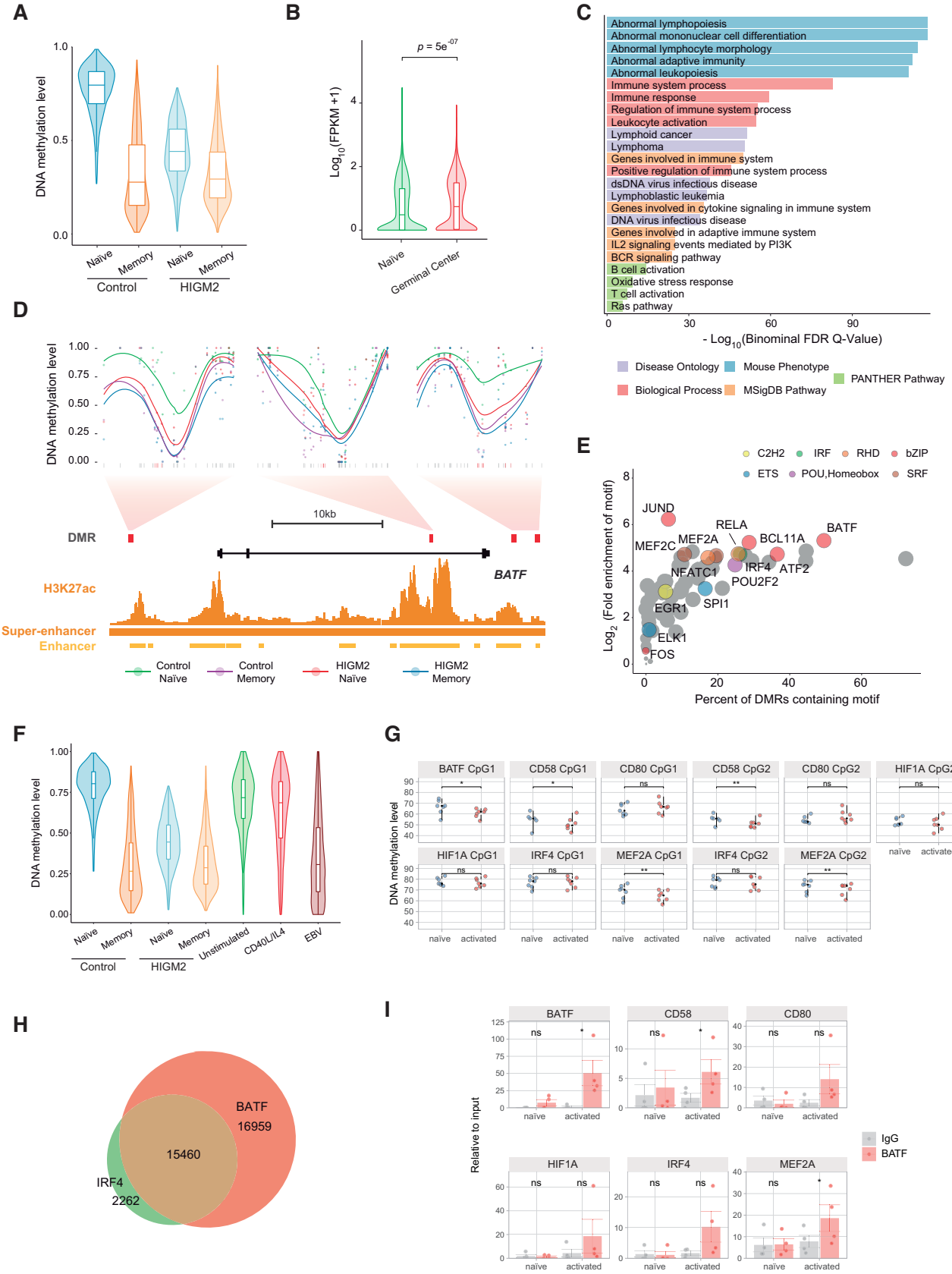


Figure 4. Role of AID in early B cell development. (**A**) Box and violin plot representation of DNA methylation of hypomethylated DMRs identified comparing control and patient naïve B cells. (**B**) Box plot showing the distribution of expression values of DMR-associated genes in naïve (green) and GC

pected, we confirmed that regions with BATF and IRF4 binding had lower DNA methylation levels in naïve B cells of HIGM2 patients compared to controls (Supplementary Figure S5C). This is also confirmed for regions that contained JUND binding, another TF downstream of BCR (Supplementary Figure S5D). This did not occur in regions enriched for other B cell-intrinsic TF binding motifs, such as RAD21, ZNF274 and MYC (Supplementary Figure S5D). Using mRNA transcription data from *IRF4* and *BATF* knockouts from GM12878 cells (55), we determined that genes with binding motifs for both TFs displayed expression changes for many of these genes associated with DMRs (Supplementary Figure S5E). Furthermore, we observed that BATF was specifically recruited to genes that displayed defective DNA demethylation in HIGM2 patients when naïve B cells were activated via BCR stimulation (Figure 4I). Altogether, these results suggest that a significant fraction of DMRs identified in HIGM2 naïve B cells may be associated with the recruitment of the BATF/IRF4 complex to these genomic sites.

The transcriptome of AID-deficient naïve B cells also reflects alterations in the BCR pathway

We then inspected whether loss of function of AID had an impact on the transcriptomic profiles of both naive and memory B cells. To this end, we performed RNA-seq of both naïve and memory B cells from the same two aforementioned HIGM2 patients. In parallel, we performed RNAseq with two healthy controls.

Comparison of the naïve B cells transcriptomes revealed statistically significant expression changes of a number of genes in HIGM2 compared to controls (DEGs: differentially expressed genes). DEGs were considered to be statistically significant when fold change ≥ 2 and adjusted P-value <0.05. 3223 genes became upregulated in HIGM2 naïve B cells in relation to healthy controls, whereas 3146 genes were downregulated. (Figure 5A and Supplementary Table S6). We then performed Gene Ontology analysis of DEGs using GREAT software which revealed enrichment of categories related to EBV activation and BCR signaling pathway (Figure 5B), consistent with our previous findings related to DNA methylation alterations in naïve B cells of HIGM2. Interestingly, we also observed differences in gene expression in memory B cells (2199 upregulated genes, 2384 downregulated) (Supplementary Figure S6A and Supplementary Table S6) which were also enriched in the aforementioned categories (Supplementary Figure S6B). Utilizing DoRothEA TF analysis, we observed that DEGs identified in naïve B cells of HIGM2 patients also displayed increased TF activity compared to controls (Figure 5C), where similar results were observed for memory B cells (Supplementary Figure S6C). Integration of the DNA methylation and expression datasets in naïve B cells did not show a clear correlation when comparing HIGM2 versus controls (Figure 5D). Interestingly, hypomethylated DMRs in HIGM2 naïve B cells showed slightly increased expression levels in HIGM2 naïve B cells in respect to control naïve B cells. The naïve B cell expression levels of hypomethylated DMRs of HIGM2 patients were more similar to those of memory B cells, reinforcing the notion of a premature activation of these cells in HIGM2 (Figure 5E).

AID deficiency causes blockade of central B cell tolerance with an expansion of pre-activated autoreactive B cells

Our results suggest that naïve B cells are pre-activated in AID deficient patients. However, the stage of B cell differentiation at which this alteration is produced remains to be established. Two independent studies reported a potential role for AID in removing autoreactive B cells during the central B cell tolerance process in the bone marrow (48,49). Specifically, immature B cells with auto-reactive BCR were activated and went into a secondary receptor editing process with an increase in AID and recombination-activating gene 2 (RAG2). If the autoreactive BCR did not lose self-antigen affinity, the genomic instability induced by the over-exposure to high levels of AID led to apoptosis. However, AID deficiency reduces the genomic damage that causes the expansion of autoreactive B cells (48,49). With that in mind, we hypothesized that the presence of naïve B cells with a pre-activation methylation signature in HIGM2 patients is a consequence of the impairment of the central B cell tolerance mechanism that causes autoreactive naïve B cells to accumulate. In fact, it has been reported that 21% of HIGM2 patients suffer some type of autoimmune disease (58).

To assess this hypothesis, we first checked whether there is an expansion of naïve autoreactive B cells in HIGM2 patients with respect to controls. To this end, we used a commercial antibody against $9G4^{+}$ IgG used to detect autoreactive clones in autoimmune diseases like systemic lupus erythematosus and rheumatoid arthritis (59,60). We observed an expansion of the naïve B cell compartment in HIGM2 patients in comparison with healthy controls (Fig-

←

(red) B cells. Gene expression data obtained from BLUEPRINT database. FPKM refers to fragments per kilobase of transcript per million mapped reads. (**C**) Results of gene set enrichment analysis using GREAT software. The plot depicts the top five enriched terms for five annotation databases ranked by P-values from the binominal distribution. All the depicted enriched terms passed a significance threshold of adjusted P-value < 0.05. (**D**) Smoothed DNA methylation data of *BATF*-associated DMRs in HIGM2 naïve B cells and H3K27ac ChIP-seq signal, super-enhancer and enhancer location in GC B cells. H3K27ac ChIP-seq data obtained from BLUEPRINT database. (**E**) Bubble scatter plot of transcription factor ChIP-seq peaks determined in GM12878 lymphoblastoid cell line in regions that correspond to DMRs identified in HIGM2 naïve B cells. Transcription factors downstream of BCR signaling are colored according to the transcription factor family. Bubble size corresponds to the logarithm of adjusted P-values. ChIP-seq data were obtained from ENCODE consortium (31). FDR refers to false discovery rate. (**F**) Box and violin plots summarizing the distribution of DNA methylation of regions that correspond to HIGM2-identified DMRs in resting B cells (Unstimulated), B cells activated with CD40L/IL4 and B cells infected with Epstein-Barr virus (EBV). Data of resting and activated B cells obtained from Hansen *et al*. (36) (**G**) Dot plot showing the DNA methylation values determined by pyrosequencing of naïve B cells (naïve) and B cells activated with anti-BCR/CD40L/IL-21 (activated) from healthy donors. (**H**) Venn diagram representing the overlap between BATF and IRF4 ChIP-seq peaks in the GM12878 cell line. (**I**) Dot and bar plot representation of BATF ChIP-qPCR results of naïve B cells (naïve) and naïve B cells activated with anti-BCR/CD40L/IL-21 (activated) from healthy donors. Statistical tests: Student's t-test (G, I) (* P-value < 0.05, ** P-value < 0.01, ns is not significant).

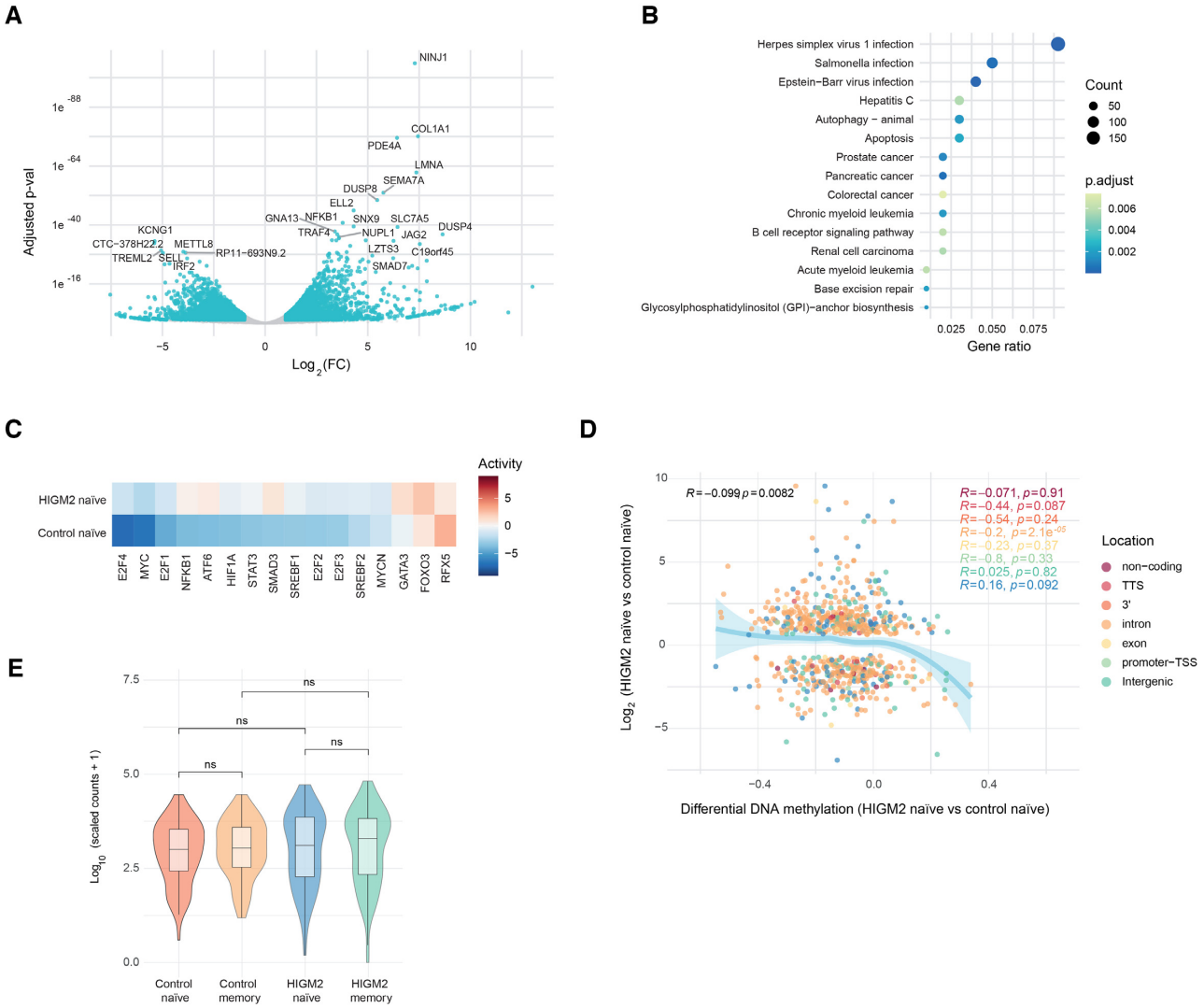


Figure 5. RNA sequencing analysis of naïve B cells transcriptomes from HIGM2 patients and healthy donors. (**A**) Volcano plot representation of the transcriptomic comparison of HIGM2 with control naïve B cells. In blue, genes that are differentially expressed (absolute of fold change ≥2 and adjusted *P*-value < 0.05). (**B**) Gene Ontology analysis representation of differentially expressed genes. All represented categories are statistically significantly enriched (adjusted *P*-value < 0.05). Gene ratio refers to the ratio between the unique genes present in the DMRs that are associated with a Gene Ontology term and all the genes potentially included in that ontology gene set. Dot size represents the number of times each term is represented in the DMRs. Dot color represents adjusted *P*-value. (**C**) Heatmap representation showing TF activity predicted using DoRothEA from mRNA expression of differentially expressed genes in control and HIGM2 naïve B cells. Color represents normalized enrichment scores (NES) where blue and red represent a decreased and increased activity, respectively, in comparison to background. (**D**) Correlation between DNA methylation differences and gene expression scaled counts ratio in naïve B cells from HIGM2 and controls. Colors represent genetic context location. (**E**) Box and violin plot representation of expression values per sample group from the genes associated with hypomethylated DMRs in HIGM2 naïve B cells. Statistical tests: Spearman's rank correlation test (D) and two-tailed unpaired Wilcoxon's (E) (ns is non significant).

ure 6A, B). We did not find an expansion of 9G4$^+$ cells in HIGM2 patients with respect to controls (Figure 6C). However, we observed a significant increase of mean fluorescence intensity for 9g4 staining (Figure 6A, D), as well as, an expansion of high 9g4$^+$ naïve B cells (Figure 6E). Next, we determined the methylation status of a selection of genes by pyrosequencing of 9G4$^-$ (non-autoreactive) and 9G4$^+$ (autoreactive) naïve B cells and found that autoreactive B cells had lower levels of DNA methylation than their non-autoreactive counterparts (Figure 6F). Overall, our results suggest that the demethylation in naïve B cells of HIGM2 patients compared with control donors is associated with an expansion of pre-activated autoreactive naïve B cells as a consequence of central B cell tolerance impairment mediated by AID deficiency.

DISCUSSION

Our results show that AID deficiency in HIGM2 syndrome results in the acquisition of aberrant DNA methylation profiles in naïve and memory B cells. First, we observed the presence of HIGM2-associated DNA methylation alterations occurring in the transition from naïve to memory B cell in which the majority are due to DNA replication-

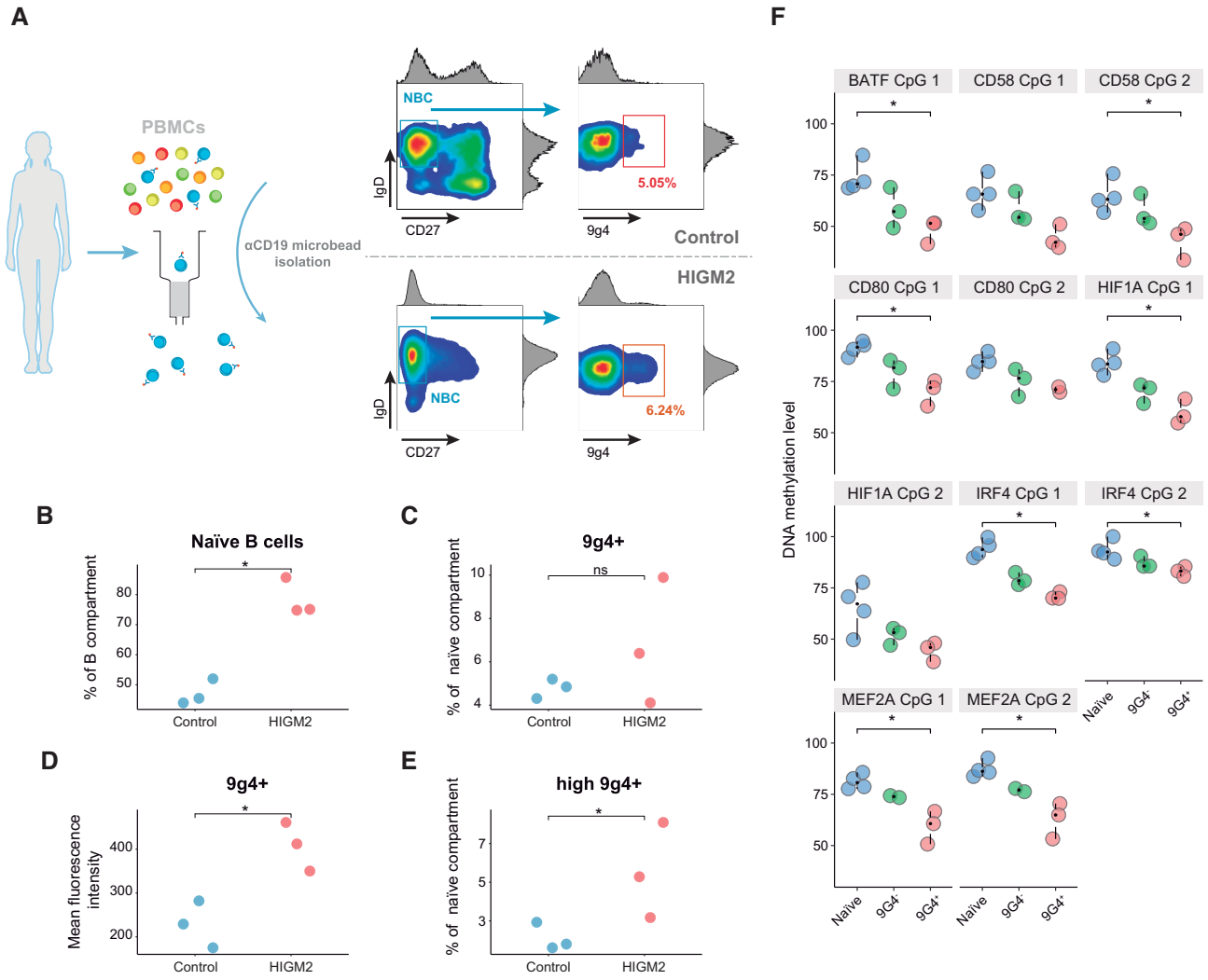


Figure 6. Central B cell tolerance impairment by AID deficiency. (**A**) Representative example of strategy for sorting naïve autoreactive B cells. (**B**) Dot plot showing the percentage of naïve B cells in the B cell compartment of HIGM2 patients and controls. (**C**) Dot plot showing the percentage of autoreactive naïve B cells in the naïve B cell compartment of HIGM2 patients and controls. (**D**) Dot plot showing the mean fluorescence of $9g4^+$ gate. (**E**) Dot plot showing the percentage of high $9g4^+$ naïve B cells in the naïve B cell compartment of HIGM2 patients and controls. (**F**) Dot plot showing the DNA methylation values determined by pyrosequencing of control naïve B cells (naïve), HIGM2 patient $9g4^-$ naïve B cells ($9g4^-$) and HIGM2 patient $9g4^+$ naïve B cells ($9g4^+$). Statistical tests: student t-test (* *P*-value < 0.05, ns is not significant).

mediated mechanisms. Second, compared to AID mouse models, we observed that AID may not be directly involved in the small proportion of those DNA methylation alterations attributable to active demethylation. Third, the comparison of naïve B cells in HIGM2 and healthy controls showed premature demethylation of genes downstream of the BCR in AID-deficient individuals, which was associated with the expansion of autoreactive B cell clones prior to the GC reaction. This reinforces a novel role for AID in preventing the expansion of autoreactive B cell clones, affecting the DNA methylation profiles of naïve B cells.

Our study indicates that AID does not play a direct role in mediating active demethylation in the transition from naïve to memory B cells. This transition is associated with the highest proportion of DNA methylation changes of the entire B cell differentiation process (20,21) and also coincides with the highest peak of AID expression. Previous studies addressing the potential participation of AID in demethylation had not considered late-replicating domains or the relationship between DNA methylation changes and the genomic features of AID targets. In our study, we determined that most of the changes taking place during the transition from naïve to memory B cells occur through passive demethylation. It is possible that memory B cells of healthy controls undergo more divisions than HIGM2 patients, similarly to what has been described for *Aicda*$^{-/-}$ mice (61). This could explain the partial impairment of passive demethylation in HIGM2 patients. For active changes, no associations with AID targets were found. These findings are also in line with those by Álvarez-Prado *et al.* (35), which indicated that AID-mediated mutation frequencies are too low. Such low frequency would be unlikely to produce a perceptible effect at the level of DNA methylation. In relation to the results obtained by Dominguez and col-

leagues (13), it is remarkable that even without the removal of PMDs in their analysis, the overlap between the DNA methylation changes and AID-associated dsDNA breaks was very small, which again reinforces the notion that AID may not be directly involved in DNA demethylation events of B cell maturation. A limitation of our analysis was the lack of effective human models to study the direct contribution of AID activity to active demethylation. Therefore, we cannot discard the possibility that the use of mouse models may not be directly attributable to the functions of human AID in the context of HIGM2. However, comparing *Aicda*-/- mouse model with a mouse model deficient for BER and MMR pathways (*Ung*-/- *Msh2*-/-), we were able to deduce that DNA methylation during B cell maturation, at least in mice, was not dependent on AID enzymatic activity. Furthermore, there is a 92% identity between mouse and human AID protein sequences, in which the catalytic and hotspot recognition domains are conserved.

Another limitation of our analysis is that we work under the assumption that all deaminase activity is attributed to AID and that no other targets exist apart from AID off-targets and Ig genes. Without further evidence, we cannot effectively conclude that AID enzymatic activity is not essential to mediate DNA demethylation during B cell maturation. Moreover, AID hotspots were originally identified as genomic locations of SHM (29), but further experiments would be required to verify that they are direct targets of AID deaminase activity in the context of human B cells.

A second major conclusion of our study concerns the identification of DNA methylation defects in naïve B cells from HIGM2 patients in relation to healthy controls. Customarily, AID expression has been regarded as being restricted to GC B cells, but some evidence suggests that AID may also have a role in central B cell tolerance (48,49,62). Along this line, during B cell development, these cells not only become activated in the GC but also in previous stages of differentiation in the bone marrow, where self-reactive immature B cells are activated in a process characterized by the upregulation of both AID and recombination-activating gene 2 (RAG2) and the downregulation of the anti-apoptotic MCL-1 (63). In this context, AID activity increases the probability of genomic damage with the subsequent activation of apoptosis through p53, which is also enhanced by the inhibition of the anti-apoptotic proteins BCL2 and MCL-1 (49). At this point, self-reactive immature B cells that are unable to correct their affinity for self-antigens by receptor editing are eliminated. In patients with AID deficiency, this mechanism of cell removal is impaired and autoreactive cells accumulate (49). Indeed, we noted that HIGM2 patients accumulated more autoreactive B cells than healthy donors, a finding that is compatible with the previously described high frequency of autoimmune disorders in this type of patients (58). The failure in AID function in these patients could be responsible for the smaller degree of genomic damage that promotes the expansion of autoreactive naïve B cells. These self-reactive B cells, owing to the persistent activation of their BCR during negative selection in the bone marrow, display a more demethylated profile in genes downstream of the BCR compared with non-autoreactive naïve B cells. However, we have only replicated this data in a selected number of CpGs and cannot assume that the pre-activation of autoreactive naïve B cells in HIGM2 accounts for all the observed alterations in genome-wide DNA methylation. It is possible that other mechanisms may actively contribute to this phenomenon. Our results, therefore, indicate that the enhanced demethylation of BCR downstream targets in HIGM2 naïve B cells may be the result of the expansion of autoreactive B cell clones as a consequence of the absence of AID. Furthermore, in order to generalize our findings to all variants of HIGM2 it would be important to extend these analyses to a bigger cohort of patients with different genetic mutations of the *AICDA* gene (38), as they can have a different impact on B cell function.

DATA AVAILABILITY

DNA methylation and expression data for this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible through GEO Series accession number GSE172012 (SuperSeries).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank CERCA Programme/Generalitat de Catalunya and the Josep Carreras Foundation for institutional support. This study makes use of data generated by the BLUEPRINT Consortium (Funding for the BLUEPRINT project provided by the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no 282510—BLUEPRINT). A full list of the investigators who contributed to the generation of the data is available from www.blueprint-epigenome.eu. We are also very grateful to Dr Javier di Noia, Dr Almudena Ramiro, Dr Javier Carmona and Dr Biola M. Javierre for useful feedback.

FUNDING

Spanish Ministry of Science, Innovation and Universities [SAF2017-88086-R to E.B.]; cofunded by FEDER funds/European Regional Development Fund (ERDF)—a way to build Europe. E.B is supported by Instituto de Salud Carlos III (ISCIII), Ref. AC18/00057, associated with i-PAD project (ERARE European Union program); P.L. and C.P. are supported by the German Cancer Aid project CO-CLL [70113869]; B.G. is funded by the Deutsche Forschungsgemeinschaft [GR1617/14-1/iPAD, SFB1160/2_B5, RESIST–EXC 2155–Project ID 390874280, CIBSS–EXC-2189–Project ID 390939984]; BMBF [GAIN 01GM1910A]. Funding for open access charge: Spanish Ministry of Science, Innovation and Universities [SAF2017-88086-R].
Conflict of interest statement. None declared.

REFERENCES

1. Revy,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O., Catalan,N., Forveille,M., Dufourcq-Lagelouse,R., Gennery,A. *et al.*

(2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*, **102**, 565–575.
2. Honjo,T., Kinoshita,K. and Muramatsu,M. (2002) Molecular mechanism of class switch recombination: linkage with somatic hypermutation. *Annu. Rev. Immunol.*, **20**, 165–196.
3. Muramatsu,M., Kinoshita,K., Fagarasan,S., Yamada,S., Shinkai,Y. and Honjo,T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, **102**, 553–563.
4. Arakawa,H., HauschiLd,J. and Buerstedde,J.M. (2002) Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science*, **295**, 1301–1306.
5. Morgan,H.D., Dean,W., Coker,H.A., Reik,W. and Petersen-Mahrt,S.K. (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming. *J. Biol. Chem.*, **279**, 52353–52360.
6. Rai,K., Huggins,I.J., James,S.R., Karpf,A.R., Jones,D.A. and Cairns,B.R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and Gadd45. *Cell*, **135**, 1201–1212.
7. Bhutani,N., Brady,J.J., Damian,M., Sacco,A., Corbel,S.Y. and Blau,H.M. (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, **463**, 1042–1047.
8. Sabag,O., Zamir,A., Keshet,I., Hecht,M., Ludwig,G., Tabib,A., Moss,J. and Cedar,H. (2014) Establishment of methylation patterns in ES cells. *Nat. Struct. Mol. Biol.*, **21**, 110–112.
9. Muramatsu,M., Sankaranand,V.S., Anant,S., Sugai,M., Kinoshita,K., Davidson,N.O. and Honjo,T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.*, **274**, 18470–18476.
10. Ramiro,A.R. and Barreto,V.M. (2015) Activation-induced cytidine deaminase and active cytidine demethylation. *Trends Biochem. Sci.*, **40**, 172–181.
11. Fritz,E.L., Rosenberg,B.R., Lay,K., Mihailović,A., Tuschl,T. and Papavasiliou,F.N. (2013) A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat. Immunol.*, **14**, 749–755.
12. Hogenbirk,M.A., Heideman,M.R., Velds,A., van den Berk,P.C.M., Kerkhoven,R.M., van Steensel,B. and Jacobs,H. (2013) Differential programming of B cells in AID deficient mice. *PLoS One*, **8**, e69815.
13. Dominguez,P.M., Teater,M., Chambwe,N., Kormaksson,M., Redmond,D., Ishii,J., Vuong,B., Chaudhuri,J., Melnick,A., Vasanthakumar,A. *et al.* (2015) DNA Methylation dynamics of germinal center B cells are mediated by AID. *Cell Rep.*, **12**, 2086–2098.
14. Abdouni,H., King,J.J., Suliman,M., Quinlan,M., Fifield,H. and Larijani,M. (2013) Zebrafish AID is capable of deaminating methylated deoxycytidines. *Nucleic Acids Res.*, **41**, 5457–5468.
15. Wijesinghe,P. and Bhagwat,A.S. (2012) Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res.*, **40**, 9206–9217.
16. Larijani,M., Frieder,D., Sonbuchner,T.M., Bransteitter,R., Goodman,M.F., Bouhassira,E.E., Scharff,M.D. and Martin,A. (2005) Methylation protects cytidines from AID-mediated deamination. *Mol. Immunol.*, **42**, 599–604.
17. Nabel,C.S., Jia,H., Ye,Y., Shen,L., Goldschmidt,H.L., Stivers,J.T., Zhang,Y. and Kohli,R.M. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.*, **8**, 751–758.
18. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (80-.).*, **324**, 930–935.
19. Ito,S., Dalessio,A.C., Taranova,O. V., Hong,K., Sowers,L.C. and Zhang,Y. (2010) Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
20. Oakes,C.C., Seifert,M., Assenov,Y., Gu,L., Przekopowitz,M., Ruppert,A.S., Wang,Q., Imbusch,C.D., Serva,A., Koser,S.D. *et al.* (2016) DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.*, **48**, 253–264.
21. Kulis,M., Merkel,A., Heath,S., Queirós,A.C., Schuyler,R.P., Castellano,G., Beekman,R., Raineri,E., Esteve,A., Clot,G. *et al.* (2015) Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.*, **47**, 746–756.
22. Victora,G.D. and Nussenzweig,M.C. (2012) Germinal Centers. *Annu. Rev. Immunol.*, **12**, 117–139.
23. Storb,U. (2014) Why does somatic hypermutation by AID require transcription of its target genes? In: *Advances in Immunology*. Academic Press Inc., Vol. **122**, pp. 253–277.
24. Chiarle,R., Zhang,Y., Frock,R.L., Lewis,S.M., Molinie,B., Ho,Y.J., Myers,D.R., Choi,V.W., Compagno,M., Malkin,D.J. *et al.* (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell*, **147**, 107–119.
25. Klein,I.A., Resch,W., Jankovic,M., Oliveira,T., Yamane,A., Nakahashi,H., Di Virgilio,M., Bothmer,A., Nussenzweig,A., Robbiani,D.F. *et al.* (2011) Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell*, **147**, 95–106.
26. Seidel,M.G., Kindle,G., Gathmann,B., Quinti,I., Buckland,M., van Montfrans,J., Scheible,R., Rusch,S., Gasteiger,L.M., Grimbacher,B. *et al.* (2019) The European Society for Immunodeficiencies (ESID) registry working definitions for the clinical diagnosis of inborn errors of immunity. *J. Allergy Clin. Immunol. Pract.*, **7**, 1763–1770.
27. Wang,Q., Gu,L., Adey,A., Radlwimmer,B., Wang,W., Hovestadt,V., Bähr,M., Wolf,S., Shendure,J., Eils,R. *et al.* (2013) Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.*, **8**, 2022–2032.
28. Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.
29. Liu,H., Liu,X., Zhang,S., Lv,J., Li,S., Shang,S., Jia,S., Wei,Y., Wang,F., Su,J. *et al.* (2016) Systematic identification and annotation of human methylation marks based on bisulfite sequencing methylomes reveals distinct roles of cell type-specific hypomethylation in the regulation of cell identity genes. *Nucleic Acids Res.*, **44**, 75–94.
30. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
31. Feingold,E.A., Good,P.J., Guyer,M.S., Kamholz,S., Liefer,L., Wetterstrand,K., Collins,F.S., Gingeras,T.R., Kampa,D., Sekinger,E.A. *et al.* (2004) The ENCODE (ENCyclopedia of DNA Elements) Project. *Science*, **306**, 636–640.
32. Kandasamy,K., Sujatha Mohan,S., Raju,R., Keerthikumar,S., Sameer Kumar,G.S., Venugopal,A.K., Telikicherla,D., Navarro,D.J., Mathivanan,S., Pecquet,C. *et al.* (2010) NetPath: a public resource of curated signal transduction pathways. *Genome Biol.*, **11**, R3.
33. Pageaud,Y., Plass,C. and Assenov,Y. (2018) Enrichment analysis with EpiAnnotator. *Bioinformatics*, **34**, 1781–1783.
34. McLean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
35. Álvarez-Prado,Á.F., Pérez-Durán,P., Pérez-García,A., Benguria,A., Torroja,C., de Yébenes,V.G. and Ramiro,A.R. (2018) A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J. Exp. Med.*, **215**, 761–771.
36. Hansen,K.D., Sabunciyan,S., Langmead,B., Nagy,N., Curley,R., Klein,G., Klein,E., Salamon,D. and Feinberg,A.P. (2014) Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res.*, **24**, 177–184.
37. Garcia-Alonso,L., Iorio,F., Matchan,A., Fonseca,N., Jaaks,P., Peat,G., Pignatelli,M., Falcone,F., Benes,C.H., Dunham,I. *et al.* (2018) Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Res.*, **78**, 769–780.
38. Durandy,A., Peron,S., Taubenheim,N. and Fischer,A. (2006) Activation-induced cytidine deaminase: structure-function relationship as based on the study of mutants. *Hum. Mutat.*, **27**, 1185–1191.
39. Hudson,T.J., Anderson,W., Aretz,A., Barker,A.D., Bell,C., Bernabé,R.R., Bhan,M.K., Calvo,F., Eerola,I., Gerhard,D.S. *et al.*

(2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
40. Zhou,W., Dinh,H.Q., Ramjan,Z., Weisenberger,D.J., Nicolet,C.M., Shen,H., Laird,P.W. and Berman,B.P. (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.*, **50**, 591–602.
41. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.-M.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
42. Durek,P., Nordström,K., Gasparoni,G., Salhab,A., Kressler,C., de Almeida,M., Bassler,K., Ulas,T., Schmidt,F., Xiong,J. *et al.* (2016) Epigenomic profiling of human CD4+ T cells supports a linear differentiation model and highlights molecular regulators of memory development. *Immunity*, **45**, 1148–1161.
43. Wu,X. and Zhang,Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*, **18**, 517–534.
44. Rogozin,I.B. and Kolchanov,N.A. (1992) Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *BBA - Gene Struct. Expr.*, **171**, 11–18.
45. Qian,J., Wang,Q., Dose,M., Pruett,N., Kieffer-Kwon,K.R., Resch,W., Liang,G., Tang,Z., Mathé,E., Benner,C. *et al.* (2014) B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell*, **159**, 1524–1537.
46. Meng,F.L., Du,Z., Federation,A., Hu,J., Wang,Q., Kieffer-Kwon,K.R., Meyers,R.M., Amor,C., Wasserman,C.R., Neuberg,D. *et al.* (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell*, **159**, 1538–1548.
47. Lio,C.W.J., Shukla,V., Samaniego-Castruita,D., González-Avalos,E., Chakraborty,A., Yue,X., Schatz,D.G., Ay,F. and Rao,A. (2019) TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer. *Sci. Immunol.*, **4**, eaau7523.
48. Kuraoka,M., Holl,T.M., Liao,D., Womble,M., Cain,D.W., Reynolds,A.E., Kelsoe,G., Kuraoka,M., Cain,D.W., Womble,M. *et al.* (2011) Activation-induced cytidine deaminase mediates central tolerance in B cells. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11560–11565.
49. Cantaert,T., Schickel,J.-N.N., Bannock,J.M.M., Ng,Y.-S.S., Massad,C., Oe,T., Wu,R., Lavoie,A., Walter,J.E.E., Notarangelo,L.D.D. *et al.* (2015) Activation-induced cytidine deaminase expression in human B cell precursors is essential for central B cell tolerance. *Immunity*, **43**, 884–895.
50. Neri,F., Rapelli,S., Krepelova,A., Incarnato,D., Parlato,C., Basile,G., Maldotti,M., Anselmi,F. and Oliviero,S. (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, **543**, 72–77.
51. Herglotz,J., Unrau,L., Hauschildt,F., Fischer,M., Kriebitzsch,N., Alawi,M., Indenbirken,D., Spohn,M., Müller,U., Ziegler,M. *et al.* (2016) Essential control of early B-cell development by Mef2 transcription factors. *Blood*, **127**, 572–581.
52. Cahir-McFarland,E.D., Carter,K., Rosenwald,A., Giltnane,J.M., Henrickson,S.E., Staudt,L.M. and Kieff,E. (2004) Role of NF-kappa B in cell survival and transcription of latent membrane protein 1-expressing or Epstein-Barr virus latency III-infected cells. *J. Virol.*, **78**, 4108–4119.
53. Mancao,C. and Hammerschmidt,W. (2007) Epstein-Barr virus latent membrane protein 2A is a B-cell receptor mimic and essential for B-cell survival. *Blood*, **110**, 3715–3721.
54. Luo,W., Weisel,F. and Shlomchik,M.J. (2018) B cell receptor and CD40 signaling are rewired for synergistic induction of the c-Myc transcription factor in germinal center B cells. *Immunity*, **48**, 313–326.
55. Ma,Y., Walsh,M.J., Bernhardt,K., Ashbaugh,C.W., Trudeau,S.J., Ashbaugh,I.Y., Jiang,S., Jiang,C., Zhao,B., Root,D.E. *et al.* (2017) CRISPR/Cas9 screens reveal Epstein-Barr virus-transformed B cell host dependency factors. *Cell Host Microbe*, **21**, 580–591.
56. Ise,W., Kohyama,M., Schraml,B.U., Zhang,T., Schwer,B., Basu,U., Alt,F.W., Tang,J., Oltz,E.M., Murphy,T.L. *et al.* (2011) The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nat. Immunol.*, **12**, 536–543.
57. Ochiai,K., Maienschein-Cline,M., Simonetti,G., Chen,J., Rosenthal,R., Brink,R., Chong,A.S., Klein,U., Dinner,A.R., Singh,H. *et al.* (2013) Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity*, **38**, 918–929.
58. Quartier,P., Bustamante,J., Sanal,O., Plebani,A., Debré,M., Deville,A., Litzman,J., Levy,J., Fermand,J.P., Lane,P. *et al.* (2004) Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to activation-induced cytidine deaminase deficiency. *Clin. Immunol.*, **110**, 22–29.
59. Richardson,C., Chida,A.S., Adlowitz,D., Silver,L., Fox,E., Jenks,S.A., Palmer,E., Wang,Y., Heimburg-Molinaro,J., Li,Q.-Z. *et al.* (2013) Molecular basis of 9G4 B cell autoreactivity in human systemic Lupus erythematosus. *J. Immunol.*, **191**, 4926–4939.
60. Moura,R.A., de la Torre,I., Leandro,M.J., Edwards,J.C.W., Fonseca,J.E. and Cambridge,G. (2012) 9G4 expression on autoantibodies to citrullinated peptides in patients with early inflammatory arthritis and established rheumatoid arthritis. *Ann. Rheum. Dis.*, **71**, A1–A93.
61. Zaheen,A., Boulianne,B., Parsa,J.Y., Ramachandran,S., Gommerman,J.L. and Martin,A. (2009) AID constrains germinal center size by rendering B cells susceptible to apoptosis. *Blood*, **114**, 547–554.
62. Meyers,G., Ng,Y.-S., Bannock,J.M., Lavoie,A., Walter,J.E., Notarangelo,L.D., Kilic,S.S., Aksu,G., Debre,M., Rieux-Laucat,F. *et al.* (2011) Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11554–11559.
63. Nemazee,D. (2006) Receptor editing in lymphocyte development and central tolerance. *Nat. Rev. Immunol.*, **6**, 728–740.