Passive and active demethylation in immune diseases with proliferation driving global hypomethylation

*Mihoko Shimada [1, 2], Yosuke Omae [2], Yuki Hitomi [3], Yoshiko Honda [1], Tohru Kodama [1], Makoto Honda [1, 4], Katsushi Tokunaga [2] and *Taku Miyagawa [1]

[1]Department of Clinical Medical Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

[2]Genome Medical Science Project, National Institute for Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan

[3]Department of Human Genetics, National Institute for Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan

[4]Seiwa Hospital, Institute of Neuropsychiatry, Tokyo, Japan

*Address correspondence to:

Taku Miyagawa, Ph.D.

Department of Clinical Medical Sciences, Tokyo Metropolitan Institute of Medical Science,

2-1-6 Kamikitazawa, Setagaya, Tokyo156-8506, Japan

Tel: +81-3-6834-2389

E-mail: miyagawa-tk@igakuken.or.jp

Mihoko Shimada, PhD

Department of Clinical Medical Sciences, Tokyo Metropolitan Institute of Medical Science,

2-1-6 Kamikitazawa, Setagaya, Tokyo, Japan

Tel: +81-3-6834-2389

E-mail: shimada-mh@igakuken.or.jp

Abstract

Global hypomethylation is frequently reported in immune-mediated diseases; however, its extent and underlying mechanisms are largely unknown. By examining seven immune-mediated diseases, we found that most exhibited global hypomethylation, particularly at CpGs in the solo-WCGW context — sequences lacking neighboring CpGs (W = A/T) — which are prone to passive demethylation during cell division. The hypomethylated solo-WCGW CpGs were predominantly located in transcriptionally repressed regions, potentially reflecting increased immune cell proliferation. Meanwhile, the majority of CpGs demethylated in response to TNF-α and TGF-β were located in transcriptionally active regions. Focusing on hypomethylated non-solo-WCGW CpGs in active regions enabled the identification of key immune-related transcription factors, such as BATF in narcolepsy and multiple sclerosis. Unlike conventional epigenome-wide association studies analyzing all CpGs, our classification-based approach, excluding CpGs susceptible to passive demethylation, could help identify CpGs more relevant to disease pathogenesis and may offer a new standard framework for future analyses.

Introduction

DNA methylation is a key epigenetic mechanism that affects gene expression and genome stability. Several studies analyzing genome-wide methylation patterns have reported global hypomethylation in immune-mediated diseases, particularly autoimmune diseases such as systemic lupus erythematosus (SLE) [1-6], rheumatoid arthritis (RA) [7-10], systemic sclerosis [11,12], and Sjögren syndrome [13-15]. Previous studies primarily focused on hypomethylations in regulatory regions and predominantly found methylation changes in these regions, such as those of interferon (IFN)-regulated genes and genes associated with the JAK-STAT pathway [16,17]; however, they do not account for all methylation sites hypomethylated in the disease.

Global hypomethylation has also been reported in various types of cancer [18,19]. These hypomethylations are primarily found in megabase-scale regions known as partially methylated domains (PMDs), which are characterized by low gene density, low GC content, late replication timing, and localization to repressed chromatin regions associated with the nuclear lamina [19,20]. Notably, a specific local sequence context associated with hypomethylation within PMDs has been identified [19]. CpG sites are categorized into two groups: “social-” (with neighboring CpG sites) and “solo-” (without neighboring CpG sites). The sequences surrounding these CpGs can be further classified into WCGW, SCGW, and SCGS contexts, where S represents the bases C or G, and W represents the bases A or T. As cell division progresses, methylation levels decrease significantly in solo-WCGW, followed by the solo-SCGW and solo-SCGS contexts, particularly within PMDs [19,21]. In contrast, no such decline in methylation levels is observed in the social-SCGS/SCGW contexts in PMDs [19,21]. Furthermore, in contrast to PMDs, the remaining regions of the genome, referred to as highly methylated domains (HMDs), show a slight decrease in methylation levels in the solo-WCGW context [21], whereas in the other contexts, methylation levels tend to increase slightly as cell division proceeds [21]. These results suggest

that some hypomethylation, especially in the solo-WCGW context, observed in cancers arises from passive demethylation due to failures in the replication and maintenance of methylation processes during cancer cell division [19,21]. However, few data are available regarding whether such passive DNA demethylation occurs in diseases other than cancer. It has also been reported that PMDs are only present in non-tumor cells of proliferative tissues such as peripheral white blood cells and the placenta [22], and another study suggested that PMDs arise only in response to environmental stimuli such as carcinogens [23].

In this study, we initially found that most of the methylation sites associated with narcolepsy type 1 (NT1) were hypomethylated in the T cells of NT1 patients, and these hypomethylated CpGs were significantly enriched in the solo-WCGW context. Although NT1 is a hypersomnia, almost all patients carry the *HLA-DQB1*06:02* allele, and T cell involvement has been reported, suggesting that immune system abnormalities contribute to the loss of orexin (hypocretine) neurons and the onset of the disease [24]. We then explored whether a similar phenomenon occurs in other immune-mediated diseases. We found significant hypomethylation in the solo-WCGW context in all of the immune-mediated diseases examined. We therefore expanded our investigation to elucidating the underlying causes of the observed CpG hypomethylation.

Results

Hypomethylation in NT1 and study design

We first conducted an epigenome-wide association study (EWAS) of NT1 using $CD4^+$ and $CD8^+$ T cells, based on previous reports that these T cell subsets react to self-antigens of target cells in NT1 [25,26]. We found that in NT1 patients, more than 97% and 80% of NT1-associated methylation sites were hypomethylated in $CD4^+$ T cells and $CD8^+$ T cells, respectively (Table S1). Next, to determine whether hypomethylated CpGs accumulate at CpGs that undergo passive demethylation during cell division, we investigated whether these hypomethylated CpGs are particularly enriched in the solo-WCGW sequence context. In both T cell types, NT1-associated hypomethylated CpGs were significantly enriched in the solo-WCGW context ($CD4^+$: $P =$ 2.57E-179, $CD8^+$: $P =$ 3.13E-21, Fig. 1).

To determine whether a similar phenomenon occurs in other immune-mediated diseases, we selected studies from previous EWASs that included at least 100 cases and 100 controls in each group and provided comprehensive, publicly available data on disease-associated methylation sites. As shown in Table S2, we analyzed studies focusing on SLE [27], RA [28], multiple sclerosis (MS) [29], MS severity [30], atopic asthma [31], and non-atopic asthma [31]. Additionally, to investigate trends in non-immune-mediated diseases, we also included studies focused on schizophrenia [32], pulmonary arterial hypertension [33], and Alzheimer's disease [34] (Table S2).

Hypomethylation in other immune-mediated diseases and its association with solo-WCGW

First, we examined whether the differentially methylated positions (DMPs) associated with each disease were specifically hypomethylated in each immune-mediated or non-immune-mediated disease. Markedly pronounced hypomethylation was observed in MS, atopic asthma, and non-atopic asthma, with hypomethylation of >95% of DMPs (Fig. 1, Table

S3). Similarly, 74.2% and 57% of DMPs were hypomethylated in SLE and MS severity, respectively, whereas in RA, only 32.2% of DMPs were hypomethylated (Fig. 1, Table S3). In contrast, hypermethylated sites were more prevalent in all cases of the three non-immune-mediated diseases examined, and no global hypomethylation was observed (Fig. 1, Table S3).

We then explored whether these hypomethylations observed in immune-mediated diseases were particularly abundant in the solo-WCGW context. The proportion of solo-WCGW in hypomethylated CpGs was significantly higher in all of the immune-mediated diseases examined (Fig. 1, Table S3). This trend was particularly strong in autoimmune diseases and NT1 (P<1E-20). In non-immune-mediated diseases, however, only Alzheimer's disease showed a nominally significant increase in solo-WCGW, but no strong associations similar to those observed in immune-mediated diseases were found (Fig. 1, Table S3).

PMDs in non-cancerous cells and their association with solo-WCGW

Next, given that hypomethylation of solo-WCGW is particularly prominent in PMDs in cancer (Fig. 2a), we investigated whether a similar phenomenon occurs in immune-mediated diseases by first identifying PMDs in immune cell subsets. Using bisulfite sequencing data from previous studies [35], we identified PMDs in 15 non-cancerous immune cell subsets, as well as in neurons and two types of cancer cells for comparison (Table S4). The same detection algorithm used in studies reporting hypomethylation of solo-WCGW in PMDs in cancer cells was used in this case [19,21]. PMDs in immune cells, cancer cells, and neurons exhibited cell/tissue specificity (Fig. 2b, Fig. S1), and two of the cancer cell types exhibited particularly large regions detected as PMDs, whereas neurons showed broader regions of PMDs compared to immune cells (Fig. 2c, Table S5). PMDs in non-cancerous cells largely overlapped between immune cells and

neurons, with an average of 60.6% of immune cell PMDs included among neuronal PMDs (conversely, 14.5% of neuronal PMDs overlapped with those in immune cells) (Table S5). In contrast, although the two cancer cell types harbored more than five times the number of PMD regions compared to neurons, only 39.2% of the immune cell PMDs overlapped with those in the cancer cells (only 1.3% of cancer cell PMDs overlapped with those in immune cells) (Table S5). Furthermore, PMDs in cancer cells were predominantly found in transcriptionally repressed regions with heterochromatin and outside regions associated with CpG islands, whereas in non-cancerous cells, PMDs were located more frequently in regions with open chromatin, transcriptionally active regions, and CpG islands (Fig. 2d, Fig. S2). These results suggest that the properties of PMDs detected in non-cancerous immune cell subsets and those detected in cancer cells differ markedly.

We also examined whether immune-mediated diseases exhibit an enrichment of hypomethylated solo-WCGW sites within PMDs in immune cells. Except for non-atopic asthma, the odds ratios (ORs) for solo-WCGW sites in immune-cell PMDs were not higher than those for overall solo-WCGW sites, and no significant enrichment was found (Fig. 2e, Table S6).

solo-WCGW sites in immune-mediated diseases are enriched in late-replicating domains and repressed chromatin

In previous studies of cancer cells, PMDs were found predominantly in late-replicating domains, in which passive demethylation is thought to occur due to insufficient time for methylation maintenance during replication, a concept known as the re-methylation window model (Fig. 2a) [19]. Therefore, we next examined the relationship between replication timing and solo-WCGW hypomethylation observed in immune-mediated diseases. Using publicly available Repli-seq data for non-cancerous B cells, we classified each analyzed CpG as belonging to

either an early- or late-replicating region and examined whether hypomethylated solo-WCGW sites occur more frequently in late-replicating regions. In all immune-mediated diseases, the ORs for hypomethylated solo-WCGW in late-replicating regions were higher than those of hypomethylated solo-WCGW overall (Fig. 2e, Table S7). However, hypomethylated solo-WCGW in late-replicating region accounted for only 7.5–39.7% of all hypomethylated solo-WCGW in each disease (Fig. 2e). In contrast to non-cancerous cells, in cancer cells, many PMDs are present in regions with late replication timing (Fig. S3a, b).

We next examined the relationship between hypomethylated solo-WCGW and chromatin regions with a closed conformation and low transcriptional activity, such as lamina-associated domains, which a previous study suggested as a second possibility for explaining passive methylation loss during cell division [19]. This mechanism is referred to as the nuclear localization model (Fig. 2a). We integrated ATAC-seq, DNase-seq, and histone ChIP-seq data for each immune cell subset and examined whether hypomethylated solo-WCGW sites in immune-mediated diseases are more prevalent in the regions identified by these assays. Particularly in NT1, RA, and MS, hypomethylated solo-WCGW sites were significantly less prevalent in regions identified using ATAC-seq and DNase-seq and also showed a lower frequency in regions marked by histone modifications associated with transcriptional activation (Fig. S4). Focusing on solo-WCGW in other diseases, sites tended to be less frequently located in regions marked by active histone modifications compared to overall hypomethylated CpGs (Fig. S4). We then classified the CpG sites as either transcriptionally active euchromatin or other non-active regions (repressed chromatin) by integrating the results of ATAC-seq, DNase-seq, and histone modification (H3K4me3, H3K9ac, H3K27ac, H3K4me1) analyses and then investigated the relationship with hypomethylated solo-WCGW. Not only were the ORs for solo-WCGW sites in repressed chromatin higher than those for hypomethylated solo-WCGW

overall, but these sites also accounted for a larger proportion of hypomethylated solo-WCGW CpGs than regions with late replication timing (25.3−86.0%, Fig. 2e, Table S8). These results suggest that although late-replicating regions are important, passive demethylation associated with replication is particularly likely to occur in repressed chromatin, especially in immune-mediated diseases.

Regarding NT1, we previously obtained methylation data from whole blood-derived DNA using a 27K array [36]. Using these data, we determined whether the same result observed in the analysis of $CD4^+$ and $CD8^+$ T cells was reproducible. In whole blood, a trend toward hypomethylation was also observed at methylation sites associated with NT1 (Table S9). Hypomethylated sites were significantly enriched in repressed chromatin (OR = 2.81, P = 2.44E-06) and more frequently found in solo CpGs lacking surrounding CpGs (OR = 2.28, P = 5.51E-04); however, they were not particularly enriched in solo-WCGW (OR = 0.976, P = 1.00) (Table S9).

Effect of cytokines on changes in global methylation

Hypomethylation in the context of solo-WCGW is not necessarily caused solely by passive demethylation along with increased cell proliferation; for instance, in immune-mediated diseases, active demethylation may occur due to the effects of various cytokines [37]. Therefore, we explored whether TNF-α and TGF-β, two cytokines known to play significant roles in immune diseases, induce hypomethylation, and whether they affect the hypomethylation of solo-WCGW, particularly in repressed chromatin. We utilized data from previous studies in which cells were treated with the cytokines, and changes in methylation sites after treatment were compared with pre-treatment [38,39]. First, we examined the direction of methylation changes in CpG sites affected by the cytokines. With TNF-α treatment, the majority of CpG sites

(86.5%) shifted toward hypomethylation (Fig. 3a), whereas with TGF-β treatment, the majority of CpG sites (90.3%) shifted toward hypermethylation (Fig. 3b). We then examined whether hypomethylated CpGs were particularly enriched in solo-WCGW within repressed chromatin. With TNF-α treatment, although solo-WCGW sites were significantly enriched (OR = 2.00, P = 8.00E-26, Table S10), hypomethylated CpGs were predominantly found in open chromatin regions with transcriptionally active histone marks (DNase-seq: OR = 7.33, P<2.22E-308, H3K27ac: OR = 2.48, P = 9.59E-65) (Fig. 3a, Table S11). Although hypomethylated CpGs represented a small fraction of all CpGs, those associated with TGF-β treatment were found in open and transcriptionally active chromatin regions in cells treated with TGF-β (ATAC-seq: OR = 6.61, P<2.22E-308, H3K27ac: OR = 4.93, P<2.22E-308, Table S11), whereas hypomethylated CpGs were less common in the solo-WCGW context (OR = 0.105, P = 1.04E-76) (Fig. 3b, Tables S10). These results suggest that at least as pertaining to the two cytokines examined in this study, hypomethylated CpGs induced by these cytokines occur predominantly in open chromatin regions and do not sufficiently explain the solo-WCGW hypomethylation in repressed chromatin observed in immune diseases.

Subsequently, we investigated the overlap between CpGs hypomethylated by TNF-α and TGF-β and disease-associated hypomethylated CpGs. Whereas CpGs hypomethylated by TGF-β did not significantly overlap with disease-associated hypomethylated CpGs in any of the diseases examined, CpGs hypomethylated by TNF-α overlapped significantly with hypomethylated regions in SLE, MS, and MS severity (Fig. 3c, Table S12).

Characterization of hypomethylated CpGs susceptible to active demethylation

Our analysis suggested that some of the hypomethylation observed in immune-mediated diseases results from passive demethylation associated with increased cell proliferation. These

CpGs appeared to be located predominantly in repressed chromatin. On the contrary, the presence of hypomethylation in open chromatin and transcriptionally active regions suggests the involvement of other demethylation-related mechanisms. To explore potential alternative mechanisms, we focused on CpG sites that are less susceptible to passive demethylation. Previous studies have indicated that for CpGs in contexts other than solo-WCGW, demethylation occurs only in PMDs, whereas solo-WCGW undergoes demethylation even in HMDs, albeit to a lesser extent, in a cell division-dependent manner. Therefore, we focused on CpG sites outside the solo-WCGW context located in transcriptionally active regions. We then identified TFs frequently bound to these sites by analyzing their overlap with TF-binding regions from ENCODE ChIP-seq datasets (Tables S13-S20). Interestingly, in our analysis of NT1, JUNB and BATF were identified in the $CD4^+$ T cell analysis (Fig. 4a, Tables S13). Functional protein association network analysis using STRING [40] showed an interaction between JUNB and BATF, indicating that these two factors form a cluster described as "Regulation of T-helper 17 cell differentiation" (FDR = 0.0187, Fig. 4a, Table S21). On the contrary, in the analysis of NT1 $CD8^+$ T cells, only FOXJ3, which accounted for just 1.5% of disease-associated CpGs, showed an association (Table S14). Similar analyses involving other diseases detected multiple TFs whose binding sites are enriched in hypomethylated sites in the diseases (Tables S15-S20, Fig. S5). An association between JUNB and BATF was also observed in the analysis of MS severity and MS (Fig. 4b, Tables S17 and S18). GATA2 showed the strongest association in both atopic asthma and non-atopic asthma, with >23% of the hypomethylated sites in atopic asthma located at GATA2 binding sites (Fig. 4c, Tables S19 and S20). The STRING analysis of atopic and non-atopic asthma identified TFs classified into clusters, such as "Granulocyte differentiation" in atopic asthma (FDR = 2.43E-06, Fig. 4c, Table

S21) and “Helix-loop-helix protein TAL-like and eosinophil fate commitment” in non-atopic asthma (FDR = 3.72E-10, Fig. S5c, Table S21).

Because the top-associated TFs in each disease exhibited similarities within several autoimmune diseases and within two types of asthma, we conducted a clustering analysis based on the odds of detecting hypomethylated sites in active regions within the binding sites of disease-associated TFs pooled across all diseases. Based on this analysis, autoimmune diseases and the two types of asthma were divided into two major clusters, with NT1 classified in the same cluster as autoimmune diseases (Fig. 4d). This result was consistent even considering all of the TFs evaluated in the study (Fig. S6a). Although the genomic regions analyzed were limited by use of the 27K array, a similar analysis using whole blood-derived DNA from NT1 patients classified NT1 in the same cluster as autoimmune diseases (Fig. S6b).

Discussion

In this study, we identified four key characteristics related to hypomethylation in immune-mediated diseases (Fig. 5). First, a tendency toward hypomethylation was observed in most immune-mediated diseases, with disease-associated hypomethylation sites being particularly abundant in the solo-WCGW context (Fig. 1). Second, such hypomethylated solo-WCGW sites were significantly more abundant in late-replicating regions and repressed chromatin (Fig. 2). Third, cytokines induced hypomethylation of certain CpG sites located primarily in open chromatin (Fig. 3). Finally, analyses focusing on hypomethylation at sites at which passive demethylation is less likely to occur enabled the identification novel factors and pathways associated with each disease (Fig. 4).

Previous studies have shown that in cancer PMDs in particular, passive demethylation occurs at CpG sites within the solo-WCGW sequence context following cell division [19,21]. In this study, we observed a consistent trend across all immune-mediated diseases examined; however, in non-immune-mediated diseases, no association with hypomethylation of solo-WCGW was found. Just as cell proliferation is enhanced in cancer, the proliferation of immune cells, including T cells and B cells, is promoted during immune responses [41,42]. Other studies have shown that an increase in the number of eosinophils is important in the pathology of asthma, involving not only the mobilization of eosinophils from the blood but also the proliferation and differentiation of progenitor cells in the lungs [43,44]. Thus, the hypomethylation of solo-WCGW observed in the present study can be attributed to the enhanced proliferation of immune cells involved in the pathology of the diseases examined. Several mechanisms have been proposed to explain the increased susceptibility to CpG demethylation in the context of solo-WCGW. One hypothesis suggests that because DNMT1 (the enzyme responsible for maintaining DNA methylation patterns) is a processive enzyme, once it binds to a CpG it efficiently methylates

surrounding CpGs [45,46]. Another hypothesis suggests that the presence of nearby methylated CpGs increases the likelihood that repair mechanisms will address demethylation [47,48]. However, the affinity of DNMT1 for CpGs in the WCGW context remains controversial [49,50], and further research is needed to clarify this issue.

Previous studies have shown that in cancers, passive demethylation of solo-WCGW sites associated with cell division occurs primarily within PMDs [19,21]; however, this was not the case in the immune-mediated diseases examined in the present study. This discrepancy is likely attributable to fundamental differences in the characteristics of PMDs between cancer cells and non-malignant immune cells. In cancer cells, PMDs are associated with late replication timing, closed chromatin, and fewer regions with histone marks related to transcriptional activation. In contrast, in non-cancerous cells, PMDs occur in open chromatin regions, areas with active transcription, and CpG islands. This is hypothesized to occur because, in non-cancerous cells, methylation sites upstream of genes for which expression is affected by factors such as the degree of cell differentiation are detected as PMDs. In immune-mediated diseases, hypomethylated solo-WCGW is predominantly found in regions with late replication timing, supporting the re-methylation window model. However, the proportion of solo-WCGW sites within late replication timing regions was not particularly high. Instead, a greater proportion of solo-WCGW sites can be attributed to repressed chromatin (Fig. 2e). These results suggest that the nuclear localization model may also be valid, particularly in the case of immune-mediated diseases.

Previous studies have reported global hypomethylation in autoimmune diseases such as SLE, with the accumulation of hypomethylated regions in IFN-related genes [16,17]. Indeed, it is possible that immune factors such as cytokines contribute to the occurrence of hypomethylation. However, in the present study, CpGs exhibiting hypomethylation following treatment with the

cytokines TNF-α or TGF-β were located primarily in open chromatin. This is consistent with previous studies showing that upon B cell stimulation, the majority of methylation changes occur in the direction of demethylation, predominantly affecting accessible chromatin and enhancer regions [51,52]. Notably, the number of cells used to examine the effect of TGF-β in this study was small (only 3), and only two cytokines were examined. Further investigation is thus necessary; however, based on our findings, the effects of cytokines alone do not sufficiently explain hypomethylation in repressed chromatin in immune-mediated diseases.

Finally, we focused on the hypomethylation of CpGs in active regions, which are less likely to undergo passive demethylation and more likely to reflect active regulatory changes, and we found that this approach enabled the identification of additional pathways not reported in earlier EWASs. To better understand the mechanisms governing disease-related changes in DNA methylation, we propose three strategies: (1) classify methylation sites as in either hypomethylated or hypermethylated regions, (2) further distinguish passive versus active hypomethylation, and (3) focus on the simultaneous changes occurring at multiple methylation sites, such as those influenced by TFs. For example, the analysis of $CD4^+$ T cells from NT1 patients found that two TFs, BATF and JUNB, bind significantly more frequently to NT1-associated CpGs that are likely to undergo active demethylation. BATF was also identified in the analysis of MS and MS severity. BATF is a member of the activator protein-1 TF family that plays a crucial role in regulating immune responses as well as the differentiation and function of various immune cells [53]. In T cells, BATF forms a trimer with JUNB and IRF4 and functions as an essential factor in the differentiation of Th17 cells, which play a pathogenic role in autoimmune diseases [54,55]. A GWAS reported that the BATF locus is linked to brain atrophy and disease progression in MS [56]. Other studies showed that experimental autoimmune encephalomyelitis in mouse and cell models could be alleviated by BATF deficiency or

inhibition [54,57]. BATF, which is induced by cytokines such as IL-6, can bind regions of closed chromatin containing methylated CpGs and promote accessibility by recruiting CTCF near target promoters [58]. Additionally, BATF facilitates the recruitment of TET2/3 to specific enhancer regions, thereby contributing to sustained transcriptional activation through the demethylation of 5-methylcytosine [52]. Given the role of BATF in regulating chromatin accessibility and inducing demethylation, CpG hypomethylation may occur at target sites in immune diseases characterized by elevated BATF expression. Although no prior evidence suggests that BATF is involved in NT1, the results of the present study suggest that NT1 exhibits hypomethylation characteristics similar to those observed in autoimmune diseases. This in turn suggests that Th17 cells, which differentiate via the overexpression of BATF, may be involved in disease onset. Further investigations, such as measuring BATF expression levels in the early stages of NT1 onset, will be necessary to obtain a more thorough understanding.

This study has two major limitations. First, for diseases other than NT1, the methylation data were derived from whole-blood DNA, which prevented the evaluation of each immune cell subset individually. In the case of NT1, global hypomethylation in repressed chromatin was observed in both $CD4^+$ and $CD8^+$ T cells, and a similar trend was observed when assessing whole blood. Therefore, unless the proliferating cells constitute a markedly small fraction of the immune cell subset, the impact could be reasonably evaluated even using whole blood. Second, methylation levels were determined using arrays in all of the EWASs we examined. Methylation arrays, particularly the 450K array, are designed to prioritize the analysis of methylation regions upstream of genes that may affect gene expression. Therefore, obtaining deeper coverage using whole-genome bisulfite sequencing could potentially expand our understanding of methylation in regions such as repressed chromatin.

In conclusion, we found that some of the hypomethylated CpGs reported in EWASs of immune-mediated diseases are affected by enhanced proliferation of immune cells, leading to passive demethylation of repressed chromatin, particularly in the solo-WCGW context (Fig. 5). These sites exhibit random methylation loss, and by focusing on CpGs in which active demethylation occurs (thereby excluding the effects of random methylation loss), we were able to identify more disease-relevant pathways. To facilitate this approach, we publicly share annotations for EPIC array CpG sites, including their presence in immune cell-defined active/repressed regions and sequence contexts such as solo-WCGW (Table S22). The findings of this study provide the basis for a novel analytical approach for EWASs examining diseases that involve cell division, such as immune-mediated diseases and cancers.

Methods

EWAS data for CD4$^+$ and CD8$^+$ T cells from NT1

In this study, we first performed an EWAS specifically focused on NT1, which is the primary subject of our research. Briefly, 42 (discovery: 28, replication: 14) patients with NT1 and 42 (discovery: 28, replication: 14) control subjects without significant sleep-disturbing events who underwent diagnostic sleep studies (polysomnogram and multiple sleep latency test) were included. NT1 was diagnosed by sleep specialists according to the International Classification of Sleep Disorders, 3rd edition [59]. CD4$^+$ and CD8$^+$ T cells were extracted from the lymphocyte fraction, and DNA was extracted from the cells. DNA methylation levels were examined using Infinium® Methylation EPIC BeadChip (Illumina, San Diego, CA, USA). The methylation rate was calculated as the β-value, and filtering and normalization (BMIQ [60], quantile normalization, and Combat [61]) were performed. M-values were used in the case-control analysis. As there were no significant differences between cases and controls in terms of background factors such as age and sex, disease-associated differentially methylated sites were identified using a *t*-test. Details regarding the EWAS data are described elsewhere [62].

External EWAS data for immune and non-immune diseases

As significant hypomethylation of solo-WCGW was identified in the investigation of NT1, we examined whether a similar phenomenon occurs in other immune-mediated diseases. For this purpose, we focused on studies that provide comprehensive data on disease-associated methylation sites in EWASs of other immune-mediated diseases. To further enhance the reliability of the results, we selected studies with sample sizes of >100 per group (cases and controls), including research on SLE [27], RA [28], MS [29], MS severity [30], atopic asthma [31], and non-atopic asthma [31] (Table S2). Additionally, we searched for EWASs on

non-immune-mediated diseases that met similar criteria to serve as references datasets, focusing on schizophrenia [32], pulmonary arterial hypertension [33], and Alzheimer's disease [34] (Table S2).

For the disease-associated methylation sites, we used data from studies that provided comprehensive information regarding these sites. However, for the MS study, no information on comprehensive disease-associated methylation sites was available, and only raw data were provided; therefore, we conducted our own data analysis. Using β-values normalized according to the default settings of the R package minfi [63], we estimated the cellular composition ratios using the DNA Methylation Age Calculator [64]. Because significant differences in the ratios of NK cells and $CD8^+$ T cells (P = 2.21E-08 and P = 0.00702) were observed between patients and healthy controls, the case-control analysis was conducted using the estimated cellular composition ratios, age, and sex as covariates. To assess the overall impact, CpG sites exhibiting $P<0.01$ were considered disease-associated methylation sites.

The DNA methylation data obtained using the Infinium HumanMethylation27 BeadChip (Illumina) on whole blood-derived DNA from individuals with NT1 were reported in our previous study [36].

Analysis of solo-WCGW

As all EWASs analyzed in this study used Illumina methylation arrays (450K or EPIC), we summarized the sequence context surrounding each CpG site based on the flanking sequence information provided in the Manifest files supplied by Illumina using a custom Python script (CpGContextCounter.py). Following previous studies [19,21], CpG sites without other CpGs within 35 bp upstream or downstream were defined as *solo-CpGs*. CpG sites flanked by A or T (denoted as 'W') on both sides were defined as *WCGW*. We assessed whether each disease-associated hypomethylated CpG site corresponded to a solo-WCGW site. We compared

the proportion of solo-WCGW sites among hypomethylated DMPs in each disease with that among all analyzed CpGs on the corresponding methylation platform (450K or EPIC) used for each disease (Fig. S7a). Statistical analyses, including the calculation of *P*-values using Fisher's exact test and the estimation of ORs, was performed using R software.

Analysis of PMDs

Previous studies of cell division-associated hypomethylation of solo-WCGW have reported that this phenomenon is particularly pronounced within PMDs detected in cancer cells. To investigate the impact of PMDs in non-cancerous immune cells, we identified PMDs using whole-genome bisulfite sequencing data [35] for immune cells by applying the same detection method used in previous studies, MethPipe [65,66] with a window size setting of 10 kb. In addition to the 15 immune cell subsets listed in Table S4, we included neurons as a reference for non-cancerous, non-immune cells and endothelial cancer cells as a representative for cancerous cells. For each cell type, PMDs were individually delineated, and PMDs detected in each cell type were consolidated, such that any region identified as a PMD in at least one cell was considered a PMD. Based on this criterion, PMDs were compiled separately for immune cells, neurons, and cancer cells. PMD detection, calculation of the total span of detected regions, assessment of overlapping regions, and identification of EPIC/450K array probes located within the detected PMDs were performed using PMD_processing.sh (Additional_file_1), region_length_summary.py, and pairwise_overlap_length.py, respectively. Principal component analysis was performed using R software. We compared the proportion of solo-WCGW sites among hypomethylated solo-WCGW sites in PMD with all analyzed solo-WCGW sites (Fig. S7a).

Investigation of replication timing

To investigate the replication timing of each gene region in non-cancerous immune cells, we explored the ENCODE Replication Timing Series [67], focusing on *Homo sapiens*, "Blood" as the tissue type, and non-cancerous cells. Data from samples GM06990, GM12878 (adult), GM12801, and GM12812 were used in this analysis. The percentage normalized signal data in BigWig format for the G1b, G2, S1, S2, S3, and S4 phases were converted to BedGraph format and subsequently merged using merge_bedgraph.sh. G1, S1, and S2 were treated as early replication phases, whereas S3, S4, and G2 were treated as late replication phases. Replication timing signals were analyzed based on the method described by Dietzen et al. [68]. Quantile normalization of $\log_2$(early/late) was performed using the RT signal of GM06990 as a reference, with the normalize.quantiles.use.target function (preprocessCore v1.44) in R software. To reduce signal noise, LOESS smoothing was applied to each chromosome using the 'loess' function in R software, with a span of 300 kb. The smoothed values were used as the final RT signal, where positive values represented early-replicating regions, and negative values represented late-replicating regions. The same analysis was performed for four samples, and regions consistently classified as early or late in all samples were defined as early or late, respectively. Regions that did not show consistent classification across all four samples were defined as non-conserved. Finally, the CpG sites targeted by the EPIC and 450K arrays were assigned to early, late, or non-conserved replication timing regions using the annotate_probes.sh script. Based on this classification, we compared the proportion of solo-WCGW sites among hypomethylated solo-WCGW sites in late-replicating regions with all analyzed solo-WCGW sites (Fig. S7a). Statistical tests were performed, following the same procedures as described in the "Analysis of solo-WCGW" section.

Examination of regions characterized by chromatin accessibility and histone modifications

To investigate in more detail the characteristics of regions containing hypomethylated solo-WCGW CpGs, particularly those associated with chromatin accessibility and histone modifications, we utilized publicly available ATAC-seq, DNase-seq, and histone ChIP-seq datasets from the ENCODE project [67]. We used all datasets with an assay title of Histone ChIP-seq, ATAC-seq, or DNase-seq that satisfied the following conditions: Organism: *Homo sapiens*; Organ: blood; Perturbation: not perturbed. Only datasets aligned to the GRCh38 genome assembly and available in BED format (narrowPeak) were considered. Furthermore, to ensure data quality, only datasets free from any flagged issues in the ENCODE audit category were included in the analysis. The retrieved data were merged by assay type, and in the case of Histone ChIP-seq, the data were separately merged for each histone mark. Genomic regions that appeared in at least one dataset of interest were considered associated regions and included in the analysis, as implemented in the script mergeBEDbyTarget.py. The overlap between hypomethylated solo-WCGW in each disease, and the regions identified by each assay, were assessed using the script GenomicFeatureLocator.py. We compared the proportion of solo-WCGW sites among hypomethylated solo-WCGW sites in repressed region with all analyzed solo-WCGW sites (Fig. S7a).

Investigation of the effect of cytokines on DNA methylation and solo-WCGW

To investigate the impact of cytokines on DNA methylation, we targeted previous studies that compared methylation rates in cells before and after cytokine treatment. GEO datasets (GEO GSE144804, GEO GSE56621) from two previous studies were used, focusing on TGF-β[39] and TNF-α[38]. For the analysis of TNF-α, human umbilical vein endothelial cells (n = 37) were treated with 20 ng/mL of TNF-α, and the data before and after treatment (24 h) were compared.

A *t*-test was performed using the filtered and standardized β-values. Given that the sample size was sufficient, the analysis was based on the threshold $P<1.0E-04$, which is the registration requirement for the EWAS catalog [69]. The analysis of TGF-β used methylation data acquired before and after treatment (120 h) of ovarian cancer cells with 5 ng/mL of TGF-β. Both a *t*-test and Wilcoxon rank-sum test were conducted because the number of samples in this dataset was small (n = 3). CpG sites were considered associated with TGF-β treatment if they satisfied the criteria of $P<0.05$ in the *t*-test and the theoretical minimum *P*-value threshold in the rank-sum test ($P<0.1$). The associations between the cytokines and solo-WCGW and disease-associated methylation sites were summarized in the same manner as described above. Chromatin accessibility and histone marks were analyzed by setting the ENCODE organ to "endothelial cell" for the TNF-α dataset and "ovary" for the TGF-β dataset. The procedure followed the same steps as described in the ENCODE section above (Examination of regions characterized by chromatin accessibility and histone modifications), except that no ATAC-seq data for endothelial cells were available in the ENCODE database.

Definitions of active and repressed chromatin regions

In this study, genomic regions were classified as those with open chromatin and active transcription, and those that are repressed. Following previous studies, the regions were defined based on a combination of detection by ATAC-seq or DNase-seq and the presence of annotated histone marks. Specifically, regions detected by either ATAC-seq or DNase-seq and marked with at least two of three histone modifications (H3K4me3, H3K9ac, or H3K27ac) were considered active promoters, whereas those marked with histone modifications H3K4me1 and H3K27ac were considered active enhancers [70]. The union of these regions was designated as an

“active region”. Genomic regions that did not fall into any of these categories were defined as “repressed chromatin”.

Investigation of CpGs less susceptible to passive demethylation

Among disease-associated hypomethylated CpG sites, those within active chromatin regions and not in the solo-WCGW context are considered less susceptible to passive demethylation, and we therefore investigated the enrichment of TF binding sites in these sites. For the ENCODE TF-ChIP-seq analysis, we obtained all datasets that met the same conditions as those used for the ATAC-seq data (conditions regarding the species and data characteristics described in the section "Examination of regions characterized by chromatin accessibility and histone modifications"). When multiple datasets were available for the same transcription factor, the regions detected in each were merged to create a unified set. We examined the overlap between disease-associated hypomethylated non-solo-WCGW sites in active regions and TF binding sites, using non-solo-WCGW sites located in active regions on both the EPIC and 450K arrays as the background set for comparison (Fig. S7b). TFs for which the binding sites overlapped disease-associated hypomethylated sites with an OR >1 and false-discovery rate (FDR) ≤5% were considered significantly enriched. TFs with very few binding sites were excluded, as their contribution to disease pathogenesis was considered limited. Specifically, we excluded TFs whose binding sites overlapped with less than 1% of disease-associated hypomethylated CpGs in active regions, excluding those in solo-WCGW contexts. In addition, a clustering analysis was performed based on the ORs of overlaps between the TFs most strongly associated with each disease and the disease-associated hypomethylated CpGs in active regions, excluding those in solo-WCGW contexts. When >10 disease-associated TFs were identified, only the top 10

showing the highest ORs were included in the analysis. The clustering analysis was also performed using all the TFs included in the analysis.

Protein-protein interaction analysis using STRING

To explore the characteristics of TFs that significantly bind to hypomethylated sites outside the solo-WCGW context within the active regions of each disease, we performed a protein-protein interaction analysis using STRING [40]. When a large number of associated TFs met the criterion of FDR <5%, the analysis was conducted using the top 5% of all tested TFs (n = 486). The STRING analysis was performed using default settings, with the minimum required interaction score set to high confidence (0.700). We performed k-means clustering based on the number of protein-protein interaction groups.

Bioinformatics workflow and command lines

The command lines used in this study are listed in Additional file 1. The python programs are also available on GitHub (https://github.com/mihshimada/Hypomethylation_immune).

Declarations

Ethics approval and consent to participate

Written informed consent was obtained from all participants or their families, and the protocols were approved by the ethics committees of all collaborating institutes.

Availability of data and materials

NT1 EWAS data has been deposited in the DNA Data Bank of Japan (DDBJ; https://www.ddbj.nig.ac.jp/index-e.html), and is available from the European Nucleotide Archive operated by the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI; https://www.ebi.ac.uk/ena/browser/home) under the accession number PRJDB15980.

Competing interests

The authors declare no conflicts of interest.

Funding

This study was supported by Grants-in-Aid for Scientific Research (grant numbers 15H04709, 17J01616, 18K15053, 19H03588, 21H02856, and 22H03006) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. This study was also supported by the Practical Research Project for Rare/Intractable Diseases and the Strategic Research Program for Brain Sciences from the Japan Agency for Medical Research and Development (AMED).

Authors' contributions

MS and YO designed the study. MH collected NT1 samples. MS, TM, and YH carried out experiments and statistical analyses. YH, TM, MH, TK, and KT interpreted the results. MS and YO drafted the manuscript, and all authors contributed to the final version of the paper.

Acknowledgements

We thank all participants who cooperated in this study. We also thank the staff of Seiwa Hospital and Koishikawa Tokyo Hospital for their support in blood sampling.

References

1. Zhu, X. *et al.* Analysis of associations between the patterns of global DNA hypomethylation and expression of DNA methyltransferase in patients with systemic lupus erythematosus. *Int J Dermatol* **50**, 697-704 (2011).
2. Javierre, B.M. *et al.* Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* **20**, 170-9 (2010).
3. Richardson, B. *et al.* Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Rheum* **33**, 1665-73 (1990).
4. Luo, Y. *et al.* Abnormal DNA methylation in T cells from patients with subacute cutaneous lupus erythematosus. *Br J Dermatol* **159**, 827-33 (2008).
5. Balada, E. *et al.* Transcript levels of DNA methyltransferases DNMT1, DNMT3A and DNMT3B in CD4+ T cells from patients with systemic lupus erythematosus. *Immunology* **124**, 339-47 (2008).
6. Balada, E., Ordi-Ros, J., Serrano-Acedo, S., Martinez-Lostao, L. & Vilardell-Tarrés, M. Transcript overexpression of the MBD2 and MBD4 genes in CD4+ T cells from systemic lupus erythematosus patients. *J Leukoc Biol* **81**, 1609-16 (2007).
7. Nakano, K., Whitaker, J.W., Boyle, D.L., Wang, W. & Firestein, G.S. DNA methylome signature in rheumatoid arthritis. *Ann Rheum Dis* **72**, 110-7 (2013).
8. de Andres, M.C. *et al.* Assessment of global DNA methylation in peripheral blood cell subpopulations of early rheumatoid arthritis before and after methotrexate. *Arthritis Res Ther* **17**, 233 (2015).
9. Karouzakis, E., Gay, R.E., Michel, B.A., Gay, S. & Neidhart, M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum* **60**, 3613-22 (2009).
10. Neidhart, M. *et al.* Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. *Arthritis Rheum* **43**, 2634-47 (2000).
11. Lian, X. *et al.* DNA demethylation of CD40l in CD4+ T cells from women with systemic sclerosis: a possible explanation for female susceptibility. *Arthritis Rheum* **64**, 2338-45 (2012).
12. Lei, W. *et al.* Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. *Scand J Rheumatol* **38**, 369-74 (2009).
13. Imgenberg-Kreuz, J. *et al.* Genome-wide DNA methylation analysis in multiple tissues in primary Sjögren's syndrome reveals regulatory effects at interferon-induced genes. *Ann Rheum Dis* **75**, 2029-2036 (2016).
14. Thabet, Y. *et al.* Epigenetic dysregulation in salivary glands from patients with primary Sjögren's syndrome may be ascribed to infiltrating B cells. *J Autoimmun* **41**, 175-81 (2013).

15. Konsta, O.D. *et al.* Defective DNA methylation in salivary gland epithelial acini from patients with Sjögren's syndrome is associated with SSB gene expression, anti-SSB/LA detection, and lymphocyte infiltration. *J Autoimmun* **68**, 30-8 (2016).
16. Vecellio, M., Wu, H., Lu, Q. & Selmi, C. The multifaceted functional role of DNA methylation in immune-mediated rheumatic diseases. *Clin Rheumatol* **40**, 459-476 (2021).
17. Mazzone, R. *et al.* The emerging role of epigenetics in human autoimmune disorders. *Clin Epigenetics* **11**, 34 (2019).
18. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239-59 (2009).
19. Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50**, 591-602 (2018).
20. Du, Q. *et al.* Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nat Commun* **10**, 416 (2019).
21. Endicott, J.L., Nolte, P.A., Shen, H. & Laird, P.W. Cell division drives DNA methylation loss in late-replicating domains in primary human cells. *Nat Commun* **13**, 6659 (2022).
22. Aran, D., Toperoff, G., Rosenberg, M. & Hellman, A. Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet* **20**, 670-80 (2011).
23. Vandiver, A.R. *et al.* Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol* **16**, 80 (2015).
24. Miyagawa, T. & Tokunaga, K. Genetics of narcolepsy. *Hum Genome Var* **6**, 4 (2019).
25. Latorre, D. *et al.* T cells in patients with narcolepsy target self-antigens of hypocretin neurons. *Nature* **562**, 63-68 (2018).
26. Pedersen, N.W. *et al.* CD8+ T cells from patients with narcolepsy and healthy controls recognize hypocretin neuron-specific antigens. *Nat Commun* **10**, 837 (2019).
27. Imgenberg-Kreuz, J. *et al.* DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann Rheum Dis* **77**, 736-743 (2018).
28. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* **31**, 142-7 (2013).
29. Kular, L. *et al.* DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. *Nat Commun* **9**, 2397 (2018).
30. Campagna, M.P. *et al.* Whole-blood methylation signatures are associated with and accurately classify multiple sclerosis disease severity. *Clin Epigenetics* **14**, 194 (2022).
31. Hoang, T.T. *et al.* Epigenome-wide association study of DNA methylation and adult asthma in the Agricultural Lung Health Study. *Eur Respir J* **56**(2020).
32. Hannon, E. *et al.* An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol* **17**, 176 (2016).

33. Ulrich, A. *et al*. Blood DNA methylation profiling identifies cathepsin Z dysregulation in pulmonary arterial hypertension. *Nat Commun* **15**, 330 (2024).
34. Zhang, L. *et al*. Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nat Commun* **11**, 6114 (2020).
35. Loyfer, N. *et al*. A DNA methylation atlas of normal human cell types. *Nature* **613**, 355-364 (2023).
36. Shimada, M., Miyagawa, T., Toyoda, H., Tokunaga, K. & Honda, M. Epigenome-wide association study of DNA methylation in narcolepsy: an integrated genetic and epigenetic approach. *Sleep* **41**(2018).
37. Rodríguez-Ubreva, J. *et al*. Inflammatory cytokines shape a changing DNA methylome in monocytes mirroring disease activity in rheumatoid arthritis. *Ann Rheum Dis* **78**, 1505-1516 (2019).
38. Rhead, B. *et al*. Global expression and CpG methylation analysis of primary endothelial cells before and after TNFa stimulation reveals gene modules enriched in inflammatory and infectious diseases and associated DMRs. *PLoS One* **15**, e0230884 (2020).
39. Cardenas, H. *et al*. TGF-β induces global changes in DNA methylation during the epithelial-to-mesenchymal transition in ovarian cancer cells. *Epigenetics* **9**, 1461-72 (2014).
40. Szklarczyk, D. *et al*. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638-D646 (2023).
41. Kaech, S.M., Wherry, E.J. & Ahmed, R. Effector and memory T-cell differentiation: implications for vaccine development. *Nat Rev Immunol* **2**, 251-62 (2002).
42. Hardy, R.R. & Hayakawa, K. B cell development pathways. *Annu Rev Immunol* **19**, 595-621 (2001).
43. Rådinger, M. *et al*. Local proliferation and mobilization of CCR3(+) CD34(+) eosinophil-lineage-committed cells in the lung. *Immunology* **132**, 144-54 (2011).
44. Amin, K., Janson, C. & Bystrom, J. Role of Eosinophil Granulocytes in Allergic Airway Inflammation Endotypes. *Scand J Immunol* **84**, 75-85 (2016).
45. Vilkaitis, G., Suetake, I., Klimasauskas, S. & Tajima, S. Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J Biol Chem* **280**, 64-72 (2005).
46. Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* **279**, 48350-9 (2004).
47. Wang, Q. *et al*. Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. *Nat Genet* **52**, 828-839 (2020).
48. Nishiyama, A. & Nakanishi, M. Navigating the DNA methylation landscape of cancer. *Trends Genet* **37**, 1012-1027 (2021).
49. Flynn, J., Azzam, R. & Reich, N. DNA binding discrimination of the murine DNA cytosine-C5 methyltransferase. *J Mol Biol* **279**, 101-16 (1998).

50. Bashtrykov, P., Ragozin, S. & Jeltsch, A. Mechanistic details of the DNA recognition by the Dnmt1 DNA methyltransferase. *FEBS Lett* **586**, 1821-3 (2012).
51. Kieffer-Kwon, K.R. *et al.* Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* **155**, 1507-20 (2013).
52. Lio, C.J. *et al.* TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the. *Sci Immunol* **4**(2019).
53. Wang, X. *et al.* The role of BATF in immune cell differentiation and autoimmune diseases. *Biomark Res* **13**, 22 (2025).
54. Schraml, B.U. *et al.* The AP-1 transcription factor Batf controls T(H)17 differentiation. *Nature* **460**, 405-9 (2009).
55. Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151**, 289-303 (2012).
56. Akkad, D.A. *et al.* Multiple sclerosis risk loci correlate with cervical cord atrophy and may explain the course of disability. *Neurogenetics* **16**, 161-8 (2015).
57. Huang, T.Y. *et al.* Phosphoenolpyruvate regulates the Th17 transcriptional program and inhibits autoimmunity. *Cell Rep* **42**, 112205 (2023).
58. Pham, D. *et al.* Batf Pioneers the Reorganization of Chromatin in Developing Effector T Cells via Ets1-Dependent Recruitment of Ctcf. *Cell Rep* **29**, 1203-1220.e7 (2019).
59. American Academy of Sleep Medicine. International Classification of Sleep Disorders-Third Edition. (2014).
60. Teschendorff, A.E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189-96 (2013).
61. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).
62. Shimada, M. *et al.* Multiomics analysis of narcolepsy T cells: global hypomethylation in solo-WCGW motif linked to T cell proliferation. *BioRxiv* (2024).
63. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
64. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013).
65. Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **8**, e81148 (2013).
66. Decato, B.E. *et al.* Characterization of universal features of partially methylated domains across tissues and species. *Epigenetics Chromatin* **13**, 39 (2020).
67. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

68. Dietzen, M. *et al*. Replication timing alterations are associated with mutation acquisition during breast and lung cancer evolution. *Nat Commun* **15**, 6039 (2024).
69. Battram, T. *et al*. The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res* **7**, 41 (2022).
70. Narita, T. *et al*. Acetylation of histone H2B marks active enhancers and predicts CBP/p300 target genes. *Nat Genet* **55**, 679-692 (2023).

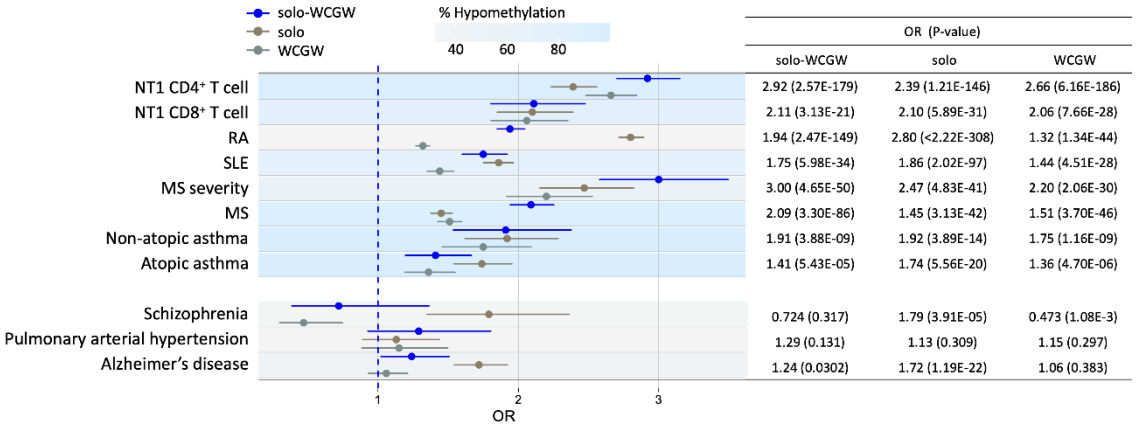

Fig. 1. Hypomethylation and the accumulation of hypomethylated CpG sites in the solo-WCGW context in immune-mediated diseases. We examined the proportion of hypomethylated disease-associated CpG sites in immune-mediated and non-immune mediated diseases. We also investigated whether these hypomethylated sites were enriched in the solo- context (i.e., lacking nearby CpGs), and specifically in the WCGW context, in which the bases flanking the CpG site are either A or T. The background of the forest plot for each disease represents the proportion of hypomethylated CpG sites relative to all disease-associated CpGs, with a stronger blue color indicating a higher degree of hypomethylation.

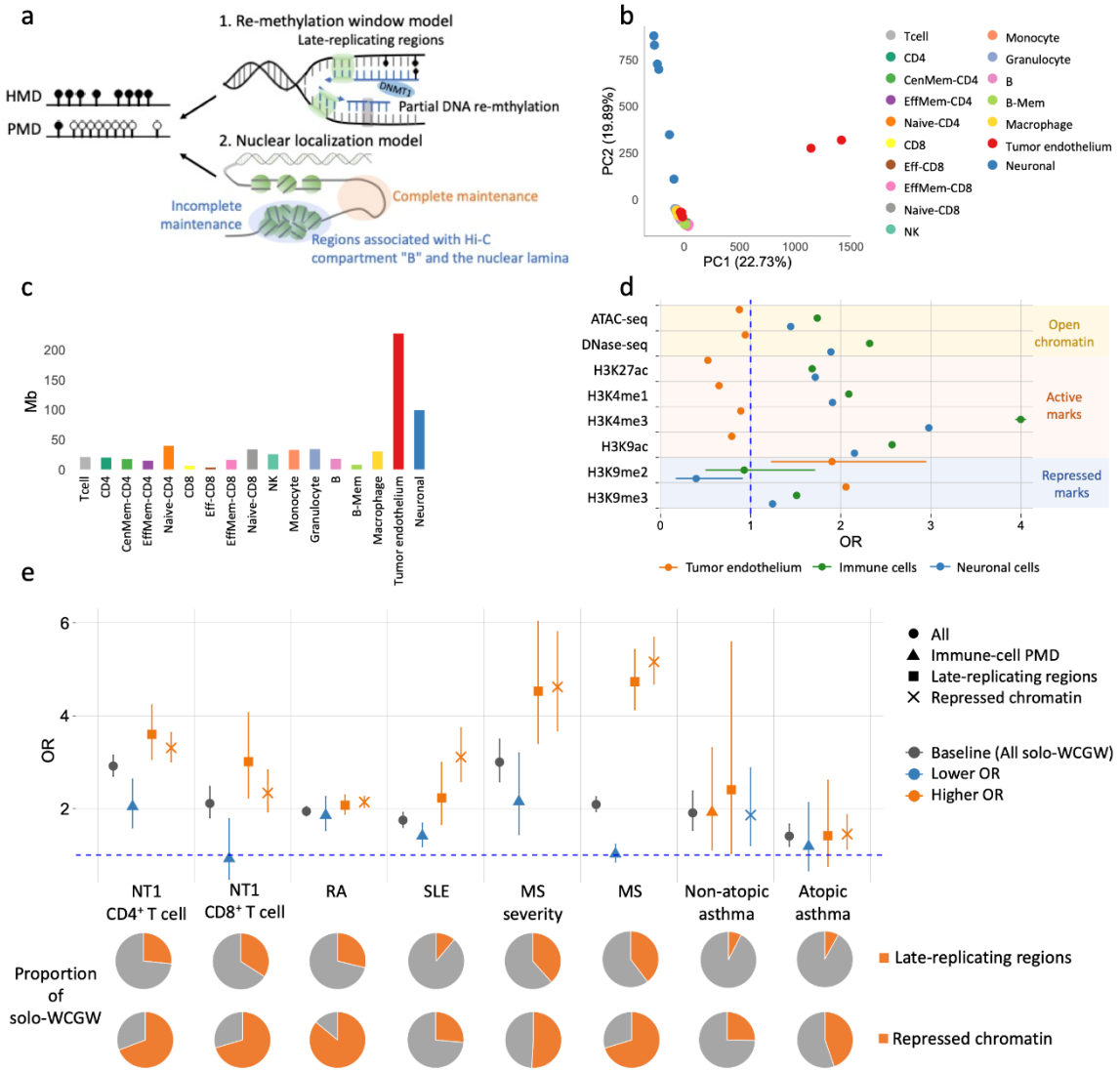

Fig. 2. Identification and characterization of regions in which passive demethylation of solo-WCGW is likely to occur. a. Mechanisms proposed in previous studies for the occurrence of passive demethylation, especially in solo-WCGW. In cancer cells, particularly within PMDs, two distinct models explaining passive demethylation have been proposed. **b.** Principal component analysis based on CpG sites located within PMDs identified in each cell type or tissue. **c.** Size (Mb) of the regions identified as PMDs in each cell/tissue. **d.** Overlap between the regions identified by ATAC-seq, DNase-seq, and histone ChIP-seq in immune cells and the PMDs in cancer, neural, and immune cells. **e.** Comparison of the odds ratios (ORs) of the presence of disease-associated solo-WCGW in immune-cell PMDs, regions with late replication timing, and repressed chromatin. For comparison of the ORs of all disease-associated solo-WCGW in Fig 1, the ORs of disease-associated solo-WCGW in each region are depicted in orange when higher and blue when lower. Pie charts show the proportion of disease-associated solo-WCGW present in late-replicating regions and repressed chromatin.

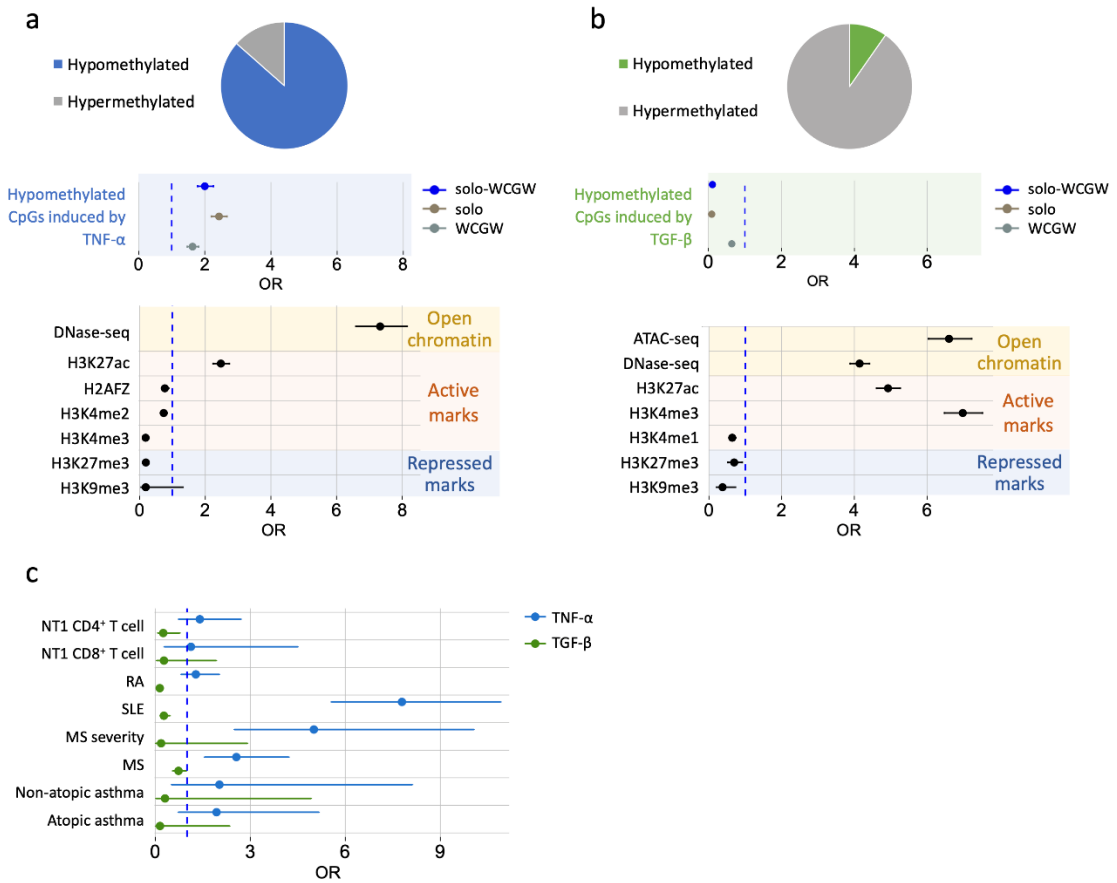

Fig. 3. Impact of cytokines on DNA methylation in active regions. a. Changes in DNA methylation upon TNF-α treatment in human umbilical vein endothelial cells. Pie chart represents the proportion of CpG sites that undergo hypomethylation among those in which methylation levels were altered by TNF-α treatment. Middle forest plot shows the ORs for solo-WCGW among hypomethylated CpG sites. Bottom forest plot shows the odds of hypomethylated CpG sites being enriched in specific genomic regions identified by DNase-seq and histone ChIP-seq analyses in endothelial cells. **b.** Changes in DNA methylation upon TGF-β treatment in ovarian cancer cell lines. Similar analyses were performed as in **a**. Bottom panel shows a forest plot based on regions identified by ATAC-seq, DNase-seq, and histone ChIP-seq in ovaries. **c.** Analysis of the odds that disease-associated hypomethylated CpG sites overlap with CpG sites hypomethylated in response to cytokine stimulation.

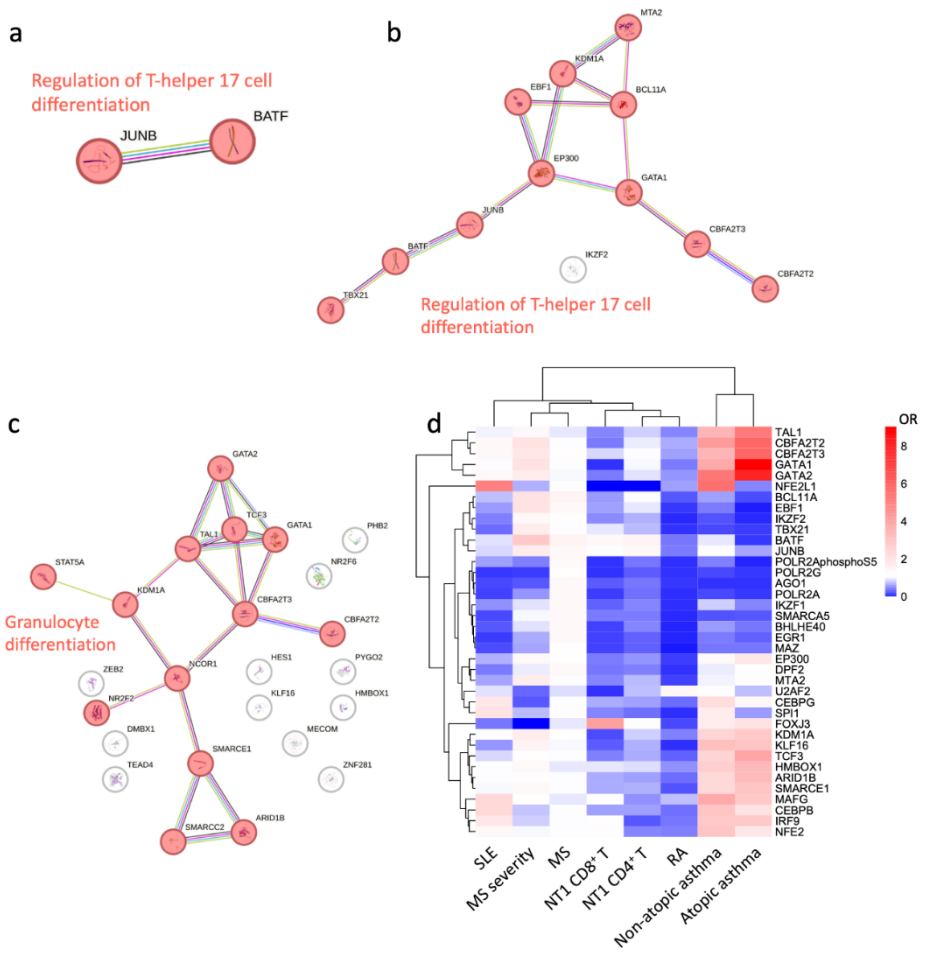

Fig. 4. Network analysis of TFs that significantly bind to CpG sites located in non-solo-WCGW contexts within active chromatin, in which active demethylation is likely to occur. **a.** Analysis of TFs in NT1 $CD4^+$ T cells. **b.** Analysis of TFs in MS severity. **c.** Analysis of TFs in atopic asthma. **d.** Heatmap and clustering analysis based on the ORs of the top-associated TFs identified in each disease.

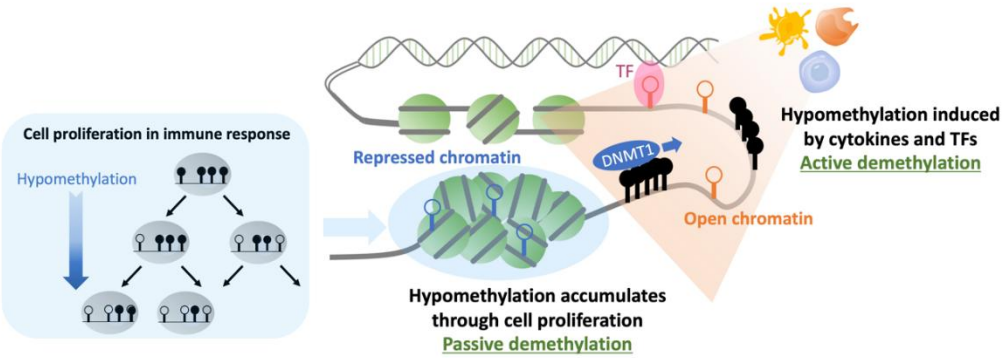

Fig. 5. Potential mechanism underlying hypomethylation in immune-mediated diseases, as suggested by the findings of this study. Based on our analysis, a tendency for hypomethylation of disease-associated methylation sites was observed in immune-mediated diseases, and some of this hypomethylation could be attributed to passive demethylation of CpG sites in repressed chromatin, particularly in the solo-WCGW context, due to enhanced cell proliferation. By contrast, in active chromatin, passive demethylation is less likely to occur, and it is suggested that active demethylation may be induced by the effects of factors such as cytokines and transcription factors (TFs).