Replication-associated solo-WCGW hypomethylation reflects cumulative immune activation across diseases

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 of Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan

[3] Department of Biometry and Genetics, Institute of Biomedical Sciences, Fukushima Medical University School of Medicine, Fukushima, Japan

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

* Correspondence authors
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, Ph.D.
Department of Clinical Medical Sciences, Tokyo Metropolitan Institute of Medical Science, 2-1-6 Kamikitazawa, Setagaya, Tokyo 156-8506, Japan
Tel: +81-3-6834-2389
E-mail: shimada-mh@igakuken.or.jp

Abstract

Global DNA hypomethylation is a hallmark of immune-mediated diseases, yet its regulatory significance remains unclear. Replication-associated loss of DNA methylation at solo-WCGW (W = A/T) CpGs has recently been proposed as a consequence of cell division. Here we systematically investigated genome-wide hypomethylation patterns across seven immune-mediated diseases. Most diseases exhibited global hypomethylation, particularly at solo-WCGW CpGs in transcriptionally repressed regions, potentially reflecting increased immune cell proliferation. By contrast, CpG sites whose methylation levels were associated with cytokine exposure or SNP genotypes were predominantly located in transcriptionally active regions. To investigate whether immunological events driving immune cell proliferation may also be imprinted in transcriptionally active regions, we searched for differentially methylated regions (DMRs) correlated with an index reflecting the extent of solo-WCGW hypomethylation. In narcolepsy, we identified a DMR within the T-cell receptor alpha (TRA) locus, and greater hypomethylation was associated with increased clonality of both TRA and TRB repertoires, with a similar pattern in an independent cohort. In multiple sclerosis, a DMR was also detected within the IGH locus encoding the B-cell receptor. Together, these findings suggest that the hypomethylation index captures the impact of disease-specific immune dynamics, while reflecting a shared epigenetic signature of immune cell proliferation across diseases.

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, hypomethylation within PMDs shows strong sequence-context dependence. CpG sites lacking neighboring CpGs ("solo" CpGs), particularly in the WCGW sequence context (W = A/T), exhibit pronounced loss of methylation with increasing cell division [19,20]. In contrast, CpG sites with neighboring CpGs ("social" CpGs) show little or no such decline [19,21]. Outside PMDs, referred to as highly methylated domains (HMDs), methylation levels also show a modest

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

71 In this study, we initially found that most of the methylation sites associated with narcolepsy
72 type 1 (NT1) were hypomethylated in the T cells of NT1 patients, and these hypomethylated
73 CpGs were significantly enriched in the solo-WCGW context. Although NT1 is clinically
74 characterized as a hypersomnia, accumulating evidence suggests that immune mechanisms
75 contribute to the loss of orexin neurons and disease onset [24]. We then explored whether a similar
76 phenomenon occurs in other immune-mediated diseases. We found significant hypomethylation
77 in the solo-WCGW context in all of the immune-mediated diseases examined. We therefore
78 expanded our investigation to elucidating the underlying causes of the observed CpG
79 hypomethylation.

Results

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

In our previous study of narcolepsy type 1 (NT1), we reported that disease-associated CpG sites were predominantly hypomethylated in $CD4^+$ and $CD8^+$ T cells [25]. 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 and found that 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 examine whether a similar pattern occurs across other immune-mediated diseases, we analyzed publicly available EWAS datasets including at least 100 cases and 100 controls per study. These datasets included studies of SLE [26], RA [27], multiple sclerosis (MS) [28], MS severity [29], atopic asthma [30], and non-atopic asthma [30] (Table S1). As a comparison, we also analyzed EWAS datasets of non-immune-mediated diseases, including schizophrenia [31], pulmonary arterial hypertension [32], and Alzheimer's disease [33] (Table S1).

We first examined whether disease-associated differentially methylated positions (DMPs) showed preferential hypomethylation. Marked global hypomethylation was observed in immune-

98 mediated diseases: more than 95.0% of DMPs were hypomethylated in MS, atopic asthma, and
99 non-atopic asthma. Similarly, 74.2% and 57.0% of DMPs were hypomethylated in SLE and MS
100 severity, respectively, whereas only 32.2% were hypomethylated in RA (Fig 1, Table S2). In
101 contrast, hypermethylated sites were more prevalent in all cases of the three non-immune-
102 mediated diseases examined, and no global hypomethylation was observed (Fig 1, Table S2).
103 We then explored whether these hypomethylations were particularly abundant in the solo-
104 WCGW context. The proportion of solo-WCGW in hypomethylated CpGs was significantly
105 higher in all of the immune-mediated diseases examined (Fig 1, Table S2). This trend was
106 particularly strong in autoimmune diseases and NT1 (P<1E-20). In non-immune-mediated
107 diseases, however, only Alzheimer's disease showed a nominally significant increase in solo-
108 WCGW, but no strong associations similar to those observed in immune-mediated diseases were
109 found (Fig 1, Table S2).
110

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

113 Although hypomethylation of solo-WCGW is particularly prominent in PMDs in cell models
114 (Fig 2a), PMDs are reported to be largely absent in non-dividing cells [23,34]. We therefore examined
115 PMDs in immune cell subsets using an established method [19,21]. We detected PMDs in 15 non-

116 cancerous immune cell subsets, as well as in neurons and tumor samples for comparison (Table
117 S3) [19,35]. The tumor samples exhibited particularly large regions detected as PMDs (Fig S1a) and
118 PMDs exhibited cell/tissue specificity (Fig S1b, S1c). Further, PMDs in tumors were
119 approximately 10 times longer than in non-cancerous cells (Fig S1d and S2). Tumor samples
120 harbored 5–10 times more PMD regions than immune cells, with only 32.7% overlapping those
121 in immune cells (Table S4). Furthermore, PMDs in tumor samples were predominantly found in
122 transcriptionally repressed regions with heterochromatin and outside of CpG island-associated
123 regions, whereas in non-cancerous cells, PMDs were located more frequently in regions with open
124 chromatin, transcriptionally active regions, and CpG islands (Fig 2b and S3). These results
125 suggest that in non-cancerous cells, local methylation variability at transcriptional regulatory
126 regions may have been identified as PMDs, indicating that PMDs with typical features such as
127 low gene density and late replication timing, as seen frequently in cancer, are largely absent in
128 non-cancerous cells. Therefore, as expected, when we examined whether the hypomethylated
129 solo-WCGW CpGs observed in immune-mediated diseases were enriched in the PMDs of
130 immune cells, except for non-atopic asthma, the odds ratios (ORs) for solo-WCGW sites in
131 immune-cell PMDs were not higher than those for overall solo-WCGW sites, and no significant
132 enrichment was found (Fig 2c, Table S5).
133

134 **Solo-WCGW sites in immune-mediated diseases are enriched**

135 **in late-replicating domains and repressed chromatin**

136 In previous *in vitro* cell model studies, PMDs were found predominantly in late-replicating

137 domains, in which passive demethylation is thought to occur during replication, a concept known

138 as the re-methylation window model (Fig 2a) [19]. We therefore next examined the relationship

139 between replication timing and solo-WCGW hypomethylation observed in immune-mediated

140 diseases. Using publicly available Repli-seq data for non-cancerous B cells, we examined whether

141 hypomethylated solo-WCGW sites are enriched in late-replicating regions. In all immune-

142 mediated diseases, the ORs for hypomethylated solo-WCGW in late-replicating regions were

143 higher than those of hypomethylated solo-WCGW overall (Fig 2c, Table S6). However,

144 hypomethylated solo-WCGW in late-replicating regions accounted for only 7.5–39.7% of all

145 hypomethylated solo-WCGW in each disease (Fig 2c).

146 We next examined the relationship between hypomethylated solo-WCGW sites and closed

147 chromatin regions such as lamina-associated domains, proposed as another explanation for

148 passive methylation loss during cell division (the nuclear localization model; Fig 2a) [19]. To assess

149 the distribution of hypomethylated solo-WCGW sites across these chromatin features, we

150 classified CpG sites as transcriptionally active euchromatin or other non-active regions (repressed

151 chromatin) using integrated ATAC-seq, DNase-seq, and histone modification data. Not only were

152 the ORs for solo-WCGW sites in repressed chromatin higher than those for hypomethylated solo-
153 WCGW overall, but these sites also accounted for a larger proportion of hypomethylated solo-
154 WCGW CpGs than regions with late replication timing (25.3−86.0%, Fig 2c, Table S7). These
155 results suggest that although late-replicating regions are important, passive demethylation
156 associated with replication is particularly likely to occur in repressed chromatin, especially in
157 immune-mediated diseases.

158 Regarding NT1, we previously obtained methylation data from whole blood-derived DNA
159 using a 27K array [36]. Using these data, we determined whether the same result observed in the
160 analysis of $CD4^+$ and $CD8^+$ T cells was reproducible. In whole blood, a trend toward
161 hypomethylation was also observed at methylation sites associated with NT1 (Table S8).
162 Hypomethylated sites were significantly enriched in repressed chromatin (OR = 2.81, P = 2.44E-
163 06) and more frequently found in solo CpGs lacking surrounding CpGs (OR = 2.28, P = 5.51E-
164 04); however, they were not particularly enriched in solo-WCGW (OR = 0.976, P = 1.00) (Table
165 S8).

166

167 Effect of cytokines on changes in global methylation

168 Hypomethylation in the context of solo-WCGW is not necessarily caused solely by passive
169 demethylation along with increased cell proliferation; for instance, in immune-mediated diseases,

170 active demethylation may occur due to the effects of various cytokines [37]. Therefore, we explored
171 whether TNF-α and TGF-β, two cytokines implicated in immune diseases, induce
172 hypomethylation, and whether they affect the hypomethylation of solo-WCGW, particularly in
173 repressed chromatin. We utilized data from previous studies in which cells were treated with the
174 cytokines, and changes in methylation sites after treatment were compared with pre-treatment [38,39].
175 First, we examined the direction of methylation changes in CpG sites affected by the cytokines.
176 With TNF-α treatment, the majority of CpG sites (86.5%) shifted toward hypomethylation (Fig
177 3a), whereas with TGF-β treatment, the majority of CpG sites (90.3%) shifted toward
178 hypermethylation (Fig 3b). We then examined whether hypomethylated CpGs were particularly
179 enriched in solo-WCGW within repressed chromatin. With TNF-α treatment, although solo-
180 WCGW sites were significantly enriched (OR = 2.00, P = 8.00E-26, Table S9), hypomethylated
181 CpGs were predominantly found in open chromatin regions with transcriptionally active histone
182 marks (DNase-seq: OR = 7.33, P<2.22E-308, H3K27ac: OR = 2.48, P = 9.59E-65) (Fig 3a, Table
183 S10). Although hypomethylated CpGs represented a small fraction of all CpGs, those associated
184 with TGF-β treatment were found in open and transcriptionally active chromatin regions in cells
185 treated with TGF-β (ATAC-seq: OR = 6.61, P<2.22E-308, H3K27ac: OR = 4.93, P<2.22E-308,
186 Table S10), whereas hypomethylated CpGs were less common in the solo-WCGW context (OR
187 = 0.105, P = 1.04E-76) (Fig 3b, Table S9). These results suggest that at least as pertaining to the

188 two cytokines examined in this study, hypomethylated CpGs induced by these cytokines occur
189 predominantly in open chromatin regions and do not sufficiently explain the solo-WCGW
190 hypomethylation in repressed chromatin observed in immune diseases.

191 Subsequently, we investigated the overlap between CpGs hypomethylated by TNF-α and TGF-
192 β and disease-associated hypomethylated CpGs. Whereas CpGs hypomethylated by TGF-β did
193 not significantly overlap with disease-associated hypomethylated CpGs in any of the diseases
194 examined, CpGs hypomethylated by TNF-α overlapped significantly with hypomethylated
195 regions in SLE, MS, and MS severity (Fig 3c, Table S11).

196

197 Transcriptionally active-region enrichment of genotype- and 198 expression-associated CpGs

199 Next, we examined CpGs associated with methylation quantitative trait loci (mQTLs) and
200 expression quantitative trait methylation (eQTMs). These CpGs, whose methylation levels are
201 influenced by genetic variants or correlate with gene expression, were analyzed for their
202 distribution between transcriptionally active and repressed regions as defined in our study. CpGs
203 located in mQTLs identified from large-scale whole-blood samples in a previous study [40] were
204 significantly enriched in transcriptionally active regions (OR = 1.52, P <2.22E-308, Fig 3d, Table
205 S12). For eQTMs, we used data registered in the eQTM-Catalog derived from two previous

206 studies in which whole blood was analyzed [41,42]. eQTM CpGs were significantly enriched in active
207 regions, and, in particular, eQTM CpGs detected in both studies exhibited a stronger enrichment
208 in active regions (OR = 2.53, *P* = 1.33E-44, Fig 3d, Table S12). Furthermore, CpGs that were
209 both influenced by SNPs and associated with gene expression—that is, CpGs overlapping mQTLs
210 and eQTMs—showed even greater enrichment in active regions (OR = 2.80, *P* = 5.21E-41, Fig
211 3d, Table S12).
212

213 Active-region DMRs associated with the global solo-WCGW 214 hypomethylation index

215 The degree of solo-WCGW hypomethylation showed strong concordance across the genome
216 within individuals (Fig S4), indicating that it can be represented by a global hypomethylation
217 index. We therefore searched for differentially methylated regions (DMRs) in transcriptionally
218 active chromatin whose methylation levels correlate with this hypomethylation index. Analyses
219 were performed in the NT1 and MS datasets, for which genome-wide single-CpG methylation
220 data were available. In the NT1 analyses, the solo-WCGW hypomethylation index correlated with
221 estimated naïve T-cell proportions (Fig S5), consistent with its association with immune cell
222 proliferation. We then searched for DMRs associated with the solo-WCGW hypomethylation
223 index after adjusting for estimated naïve T-cell proportions. No DMRs replicated at FWER<0.05

224 were detected in the $CD8^+$ T-cell analysis (Table S13, S14). In contrast, the DMR in the T cell
225 receptor alpha (TRA) locus showing the strongest association in the $CD4^+$ T-cell analysis was
226 replicated in an independent sample set (FWER<0.05) (Table1, Fig 4a, Table S15).

227 We next estimated T-cell receptor clonality using RNA-seq data derived from the same samples.
228 TCR repertoires reconstructed with TRUST4 [43] revealed that samples with higher levels of solo-
229 WCGW hypomethylation exhibited increased clonality in both the TRA (P = 0.0394) and TRB
230 repertoires (P = 0.00847) (Fig 4b). To assess whether the hypomethylation index explains
231 variation in T-cell clonality beyond known covariates, we compared regression models with
232 sequential covariate adjustment. Incorporating the hypomethylation index increased the adjusted
233 R^2 from -0.01 to 0.05 and resulted in the lowest AIC among the models tested (ΔAIC = -2.8),
234 indicating improved model fit (Fig 4c).

235 In the MS analysis, although an independent replication dataset was not available, 15 DMRs
236 were identified at FDR<0.2. Among these, a DMR located within the immunoglobulin heavy-
237 chain (IGH) locus was detected (Table S16).

238

239 Prioritizing disease-associated genes through eQTM mapping

240 Given that chromatin states may differentially influence regulatory potential and that
241 hypomethylation in repressed chromatin appears to capture a cumulative footprint of immune

242 activation, we next asked whether EWAS signals are preferentially enriched in active regulatory
243 regions. Using eQTM associations, disease-associated CpG sites were linked to gene expression,
244 only when genes were supported by multiple CpGs meeting the specified criteria (Fig S6). As a
245 result, no candidate genes were identified in repressed regions across all diseases, reinforcing the
246 limited functional impact of methylation loss in silent chromatin. In contrast, our analysis of
247 transcriptionally active regions revealed disease-associated genes in SLE and MS (Table S17,
248 Table S18, Fig S7a and S7b). The STRING analysis [44] indicated that genes negatively correlated
249 with SLE-associated hypomethylated CpG sites were significantly enriched in the type I
250 interferon (IFN-α/β) signaling pathway (Fig S7a), which had not been reported in the original
251 EWAS of SLE [26]. Among these protein-coding genes, 80.6% have been implicated in SLE in
252 large-scale whole-blood gene-expression studies [45,46] (Table S17). As illustrated by the *IFITM1*
253 example in Fig S7c, multiple CpG sites in the eQTMs correlated with *IFITM1* were associated
254 with SLE, consistent with increased *IFITM1* expression in SLE. Similarly, STRING analysis of
255 genes whose expression is negatively associated with hypomethylated CpG sites in MS
256 demonstrated significant enrichment in interferon signaling pathway (Fig S7b). Some of these
257 genes had been reported to be associated with Type I interferon therapy in MS (Table S18).

258 # Discussion

259 In this study, we identified five key characteristics related to hypomethylation in immune-
260 mediated diseases (Fig 5). First, hypomethylation was observed in most immune-mediated
261 diseases, with disease-associated hypomethylation sites being particularly abundant in the solo-
262 WCGW context (Fig 1). Second, such hypomethylated solo-WCGW sites were significantly more
263 abundant in late-replicating regions and repressed chromatin (Fig 2). Third, cytokine-induced
264 hypomethylation occurred primarily in open chromatin (Fig 3), and CpGs in mQTLs and eQTMs
265 were also enriched in transcriptionally active regions (Fig 4a). Fourth, the hypomethylation index
266 of solo-WCGW sites in repressed chromatin was associated with the *TRA* locus in NT1 and
267 correlated with T-cell clonality. In MS, the index was associated with the *BCR* locus, collectively
268 suggesting that it captures an epigenetic footprint of cumulative immune activation. Finally,
269 integrating disease-associated hypomethylated CpGs with eQTM information highlighted
270 disease-associated genes and pathways when analyses were restricted to transcriptionally active
271 regions, including the type I interferon signaling pathway in SLE (Fig S7). These results suggest
272 that hypomethylation in repressed chromatin largely reflects indirect effects of immune activation
273 through cell proliferation, whereas hypomethylation in active regulatory regions more directly
274 captures disease-associated regulatory changes (Fig 5).

275 Previous studies have shown that passive demethylation occurs at CpG sites within the solo-
276 WCGW sequence context following cell division [19,21]. In this study, we observed a consistent
277 trend across all immune-mediated diseases examined. Just as cell proliferation is enhanced in
278 cancer, the proliferation of immune cells, including T cells and B cells, is promoted during
279 immune responses [47,48]. Other studies have shown that eosinophilia in asthma involves not only
280 mobilization from the blood but also proliferation and differentiation of progenitor cells in the
281 lungs [49,50]. Thus, the hypomethylation of solo-WCGW observed in immune-mediated diseases
282 can be attributed to the enhanced proliferation of immune cells involved in the pathology. Several
283 mechanisms have been proposed to explain the increased susceptibility to CpG demethylation in
284 the context of solo-WCGW. One hypothesis suggests that because DNMT1 (the enzyme
285 responsible for maintaining DNA methylation patterns) is a processive enzyme, once it binds to
286 a CpG it efficiently methylates surrounding CpGs [51,52]. Another hypothesis suggests that the
287 presence of nearby methylated CpGs increases the likelihood that repair mechanisms will address
288 demethylation [53,54]. However, the affinity of DNMT1 for CpGs in the WCGW context remains
289 controversial [55,56], and further research is needed to clarify this issue.

290 Previous studies have shown that passive demethylation of solo-WCGW sites associated with
291 cell division occurs primarily within PMDs, especially prominent in cancerous cells [19,21]; however,
292 this was not the case in the immune-mediated diseases examined in the present study. This

293 discrepancy is likely attributable to fundamental differences in the characteristics of PMDs
294 between cancer cells and non-malignant immune cells. In cancer cells, large-scale
295 hypomethylated regions are clearly detected as PMDs. In contrast, in immune cells,
296 transcriptional regulatory regions can be detected as PMDs, suggesting that normal immune cells
297 may not harbor PMDs as prominent as those observed in cancer cells. In immune-mediated
298 diseases, hypomethylated solo-WCGW sites are predominantly found in regions with late
299 replication timing, supporting the re-methylation window model. However, because the total
300 number of analyzed CpGs located in late-replicating regions is relatively small, the overall
301 proportion of solo-WCGW sites in these regions remains limited. Instead, a greater proportion of
302 solo-WCGW sites can be attributed to repressed chromatin (Fig 2c). These results suggest that
303 the nuclear localization model may also be valid, particularly in the case of immune-mediated
304 diseases.

305 Previous studies have reported global hypomethylation in autoimmune diseases such as SLE,
306 with the accumulation of hypomethylated regions in IFN-related genes [16,17]. Indeed, it is possible
307 that immune factors such as cytokines contribute to the occurrence of hypomethylation. However,
308 in the present study, CpGs exhibiting hypomethylation following treatment with the cytokines
309 TNF-α or TGF-β were located primarily in open chromatin. This is consistent with previous
310 studies showing that upon B cell stimulation, the majority of methylation changes occur in the

311 direction of demethylation, predominantly affecting accessible chromatin and enhancer regions
312 [57,58]. Given the limited sample size for TGF-β (n = 3) and the small number of cytokines examined,
313 further investigation is warranted. Nevertheless, our findings suggest that cytokine effects alone
314 are insufficient to explain hypomethylation in repressed chromatin in immune-mediated diseases.
315 Similar to the hypomethylation induced by cytokines, our analysis revealed that CpGs in
316 mQTLs and eQTMs are enriched in transcriptionally active regions, indicating that focusing on
317 transcriptionally active regions may not only enhance the interpretation of DNA methylation and
318 gene expression analyses, but also facilitate the identification of genetic variants associated with
319 traits in genome-wide association studies. Indeed, when we used eQTM relationships to search
320 for genes whose expression is suggested to be linked to disease-associated hypomethylation,
321 genes were identified only in the analyses focusing on transcriptionally active regions in both the
322 SLE and MS datasets in which disease-associated genes were detected. In SLE we identified the
323 involvement of multiple genes associated with the well-established type I interferon pathway,
324 which was not detected in the original EWAS. In MS, interferon-related pathways were likewise
325 detected; this is likely attributable to the fact that 65% of the patients included in the MS EWAS
326 were receiving treatment, and 69.2% of those treated were on interferon therapy [59]. Focusing on
327 active chromatin regions and leveraging eQTM-based identifications for disease-associated genes
328 are therefore informative. Accordingly, in Table S19 we provide, for each CpG on the methylation

329 array, the sequence context (e.g., solo-WCGW), chromatin state annotation (active or repressed),
330 and the eQTM information detected in the study by Keshawarz *et al.* [41] (i.e., the genes with which
331 it is positively or negatively correlated, the number of CpGs correlated with that gene, and the
332 rank of the CpG in terms of association strength within that gene).

333 Finally, one of the most important observations in this study is that methylation levels at solo-
334 WCGW sites in repressed chromatin were highly consistent within individuals, allowing their
335 average level to be summarized as a hypomethylation index. We hypothesized that if immune
336 activation promotes cell proliferation, the degree of such proliferation might be reflected in this
337 index. Based on this hypothesis, we examined whether methylation changes in transcriptionally
338 active regions correlate with the hypomethylation index. In the NT1 analysis of $CD4^+$ T cells, a
339 DMR within the TRA locus was identified as significantly associated with the index, whereas in
340 the MS analysis, a DMR within the IGH locus, which encodes the B-cell receptor heavy chain,
341 showed a similar association after adjustment for cell composition. Notably, in NT1, samples with
342 lower values of the index—indicating more extensive hypomethylation—showed increased T-
343 cell receptor clonality, suggesting expansion of specific T-cell clones. Meanwhile, in B cells from
344 MS patients, hypomethylation has been reported, with lower methylation levels associated with
345 higher IgM levels [60]. This observation supports the BCR-related signal suggested by our MS
346 analysis. Together, these findings suggest that the hypomethylation index may function as an

347 epigenetic footprint of cumulative immune activation. Although variation in immune cell
348 composition may influence DNA methylation patterns, our observations were consistent across
349 multiple immune-mediated diseases and remained evident after adjustment for cell composition,
350 suggesting that the observed hypomethylation largely reflects proliferation-associated processes
351 rather than differences in cell-type proportions.

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

363 In conclusion, we found that some of the hypomethylated CpGs reported in EWASs of
364 immune-mediated diseases are affected by enhanced proliferation of immune cells, leading to

365 passive demethylation of repressed chromatin, particularly in the solo-WCGW context (Fig 5).
366 This hypomethylation may represent an epigenetic footprint of cumulative immune activation and
367 could serve as a molecular proxy for cumulative immune burden during disease development. By
368 focusing on CpGs in which active demethylation occurs (thereby excluding the effects of random
369 methylation loss), we were able to identify more disease-relevant pathways. These findings
370 provide a conceptual framework for interpreting disease-associated methylation changes and may
371 help refine EWAS strategies for diseases involving cell proliferation.

372 # Materials and Methods

373 Extended descriptions of the methods are provided in the Supplementary Methods.

374

375 ## EWAS data for immune and non-immune diseases

376 For NT1, we used an in-house dataset (see Supplementary Methods for details). After
377 identifying significant hypomethylation of solo-WCGW in NT1, we examined whether a similar
378 pattern occurs in other immune-mediated diseases. We focused on studies that provide
379 comprehensive data on disease-associated methylation sites in EWASs of other immune-mediated
380 diseases. To further enhance the reliability of the results, we selected studies with sample sizes of
381 >100 per group (cases and controls), including research on SLE [26], RA [27], MS [28], MS severity [29],
382 atopic asthma [30], and non-atopic asthma [30] (Table S1). Additionally, we analyzed EWAS datasets
383 on non-immune-mediated diseases meeting similar criteria as references, focusing on
384 schizophrenia [31], pulmonary arterial hypertension [32], and Alzheimer's disease [33] (Table S2).

385 For the disease-associated methylation sites, we used data from studies that provided
386 comprehensive information regarding these sites. However, for the MS study, no information on
387 comprehensive disease-associated methylation sites was available, and only raw data were
388 provided; therefore, we conducted our own data analysis (see Supplementary Methods for details).

389

390 # Analysis of solo-WCGW

391 As all EWASs analyzed in this study used Illumina methylation arrays (450K or EPIC), we
392 summarized the sequence context surrounding each CpG site based on the flanking sequence
393 information provided in the Manifest files supplied by Illumina using a custom Python script
394 (CpGContextCounter.py). Following previous studies [19,21], CpG sites without other CpGs within
395 35 bp upstream or downstream were defined as *solo-CpGs*. CpG sites flanked by A or T (denoted
396 as 'W') on both sides were defined as *WCGW*. We assessed whether each disease-associated
397 hypomethylated CpG site corresponded to a solo-WCGW site. We compared the proportion of
398 solo-WCGW sites among hypomethylated DMPs in each disease with that among all analyzed
399 CpGs on the corresponding methylation platform (450K or EPIC) used for each disease (Fig S8).
400 Statistical analyses, including the calculation of *P*-values using Fisher's exact test and the
401 estimation of ORs, were performed using R software.

402

403 # Analysis of PMDs

404 To investigate the impact of PMDs in non-cancerous immune cells, we identified PMDs using
405 whole-genome bisulfite sequencing data [35] from immune cells with, MethPipe [61,62]. In addition to
406 the 15 immune cell subsets listed in Table S3, we included neurons as a reference for non-
407 cancerous, non-immune cells. The nine whole-genome bisulfite sequence data of tumor samples

408 [19] were also included as a representative for cancerous cells. For each cell type, PMDs were
409 individually delineated, and PMDs detected in each cell type were consolidated, such that any
410 region identified as a PMD in at least one cell was considered a PMD. Based on this criterion,
411 PMDs were compiled separately for immune cells, neurons, and cancer cells. PMD detection,
412 calculation of the total span of detected regions, assessment of overlapping regions, and
413 identification of EPIC/450K array probes located within the detected PMDs were performed using
414 PMD_processing.sh (Additional_file_2), region_length_summary.py, and
415 pairwise_overlap_length.py, respectively. Principal component analysis was performed using R
416 software. To investigate the chromatin states enriched in PMD regions, we utilized ATAC-seq,
417 DNase-seq, and histone ChIP-seq data from the ENCODE project [63]. The detailed parameters for
418 data processing for immune cells are described in the "Examination of regions characterized by
419 chromatin accessibility and histone modifications" section. The analyses for neuron and tumor
420 samples were consistent with those for immune cells, except for the tissue types, which were
421 determined by the biosample metadata. The list of biosamples used for neuron and tumor samples
422 is provided in Table S20. Due to the lack of available data for H3K9me2 in the tumor and neuronal
423 analyses and H3K9ac in the neuron analysis, we substituted blood-derived data for these histone
424 marks. We compared the proportion of solo-WCGW sites among hypomethylated solo-WCGW
425 sites in PMD with all analyzed solo-WCGW sites (Fig S8).

426

427 ## Examination of chromatin accessibility, histone modifications,

428 ## and replication timing

429 To investigate in more detail the characteristics of regions containing hypomethylated solo-

430 WCGW CpGs, particularly those associated with chromatin accessibility and histone

431 modifications, we utilized publicly available ATAC-seq, DNase-seq, and histone ChIP-seq

432 datasets from the ENCODE project [63]. Further details, including replication timing analyses, are

433 described in the Supplementary Methods.

434

435 ## Investigation of the effect of cytokines on DNA methylation and

436 ## solo-WCGW

437 To investigate the impact of cytokines on DNA methylation, we targeted previous studies that

438 compared methylation rates in cells before and after cytokine treatment. GEO datasets (GEO

439 GSE144804, GEO GSE56621) from two previous studies were used, focusing on TGF-β[39] and

440 TNF-α[38]. For the analysis of TNF-α, human umbilical vein endothelial cells (n = 37) were treated

441 with 20 ng/mL of TNF-α, and the data before and after treatment (24 h) were compared. A *t*-test

442 was performed using the filtered and standardized β-values. Given that the sample size was

443 sufficient, the analysis was based on the threshold P<1.0E-04, which is the registration

444 requirement for the EWAS catalog [64]. The analysis of TGF-β used methylation data acquired
445 before and after treatment (120 h) of ovarian cancer cells with 5 ng/mL of TGF-β. Both a *t*-test
446 and Wilcoxon rank-sum test were conducted because the number of samples in this dataset was
447 small (n = 3). CpG sites were considered associated with TGF-β treatment if they satisfied the
448 criteria of $P<0.05$ in the *t*-test and the theoretical minimum *P*-value threshold in the rank-sum test
449 ($P<0.1$). The associations between the cytokines and solo-WCGW and disease-associated
450 methylation sites were summarized in the same manner as described above. Chromatin
451 accessibility and histone marks were analyzed by setting the ENCODE organ to "endothelial cell"
452 for the TNF-α dataset and "ovary" for the TGF-β dataset. The procedure followed the same steps
453 as described in the Supplementary Methods (Examination of regions characterized by chromatin
454 accessibility and histone modifications), except that no ATAC-seq data for endothelial cells were
455 available in the ENCODE database.
456

457 **Definitions of active and repressed chromatin regions**

458 In this study, genomic regions were classified as those with open chromatin and active
459 transcription, and those that are repressed. Following previous studies, the regions were defined
460 based on a combination of detection by ATAC-seq or DNase-seq and the presence of annotated
461 histone marks. Specifically, regions detected by either ATAC-seq or DNase-seq and marked with

462 at least two of three histone modifications (H3K4me3, H3K9ac, or H3K27ac) were considered
463 active promoters, whereas those marked with histone modifications H3K4me1 and H3K27ac
464 were considered active enhancers [65]. The union of these regions was designated as an "active
465 region". Genomic regions that did not fall into any of these categories were defined as "repressed
466 chromatin".
467

468 Distribution of CpGs in mQTLs and eQTMs across active and 469 repressed chromatin regions

470 We examined the relationship between transcriptionally active/repressed regions and CpGs
471 located in mQTLs or eQTMs. For mQTLs, we used loci identified in previous large-scale studies
472 that measured DNA methylation in whole blood using the Illumina EPIC array [40]. For eQTMs,
473 we used data from two studies registered in the eQTM-Catalog
474 (https://shiny.crc.pitt.edu/eqtm_browser/), both of which measured DNA methylation in whole
475 blood [41,42]. Although one study was profiled on the EPIC array, its sample size was <50 [42];
476 therefore, we complemented it by analyzing eQTM data identified in more than 2,000 samples
477 using the 450K array [41]. Among CpGs assayed on the EPIC array, we compared the proportion of
478 mQTL versus non-mQTL CpGs located in active regions. A similar comparison was performed
479 for eQTMs, focusing on CpG sites measurable with the Illumina 450K array. We also assessed

480 CpGs shared between mQTLs and eQTMs. For eQTMs, analyses were conducted both using the
481 union of eQTMs detected in either study and using only those detected in both studies.
482

483 Identification of DMRs associated with the hypomethylation 484 index

485 To assess whether hypomethylation at solo-WCGW sites in repressed chromatin showed
486 consistent inter-individual variation across samples, we visualized methylation levels using
487 heatmaps generated in R. Because the relative levels of hypomethylation were largely consistent
488 within individuals across these sites, we summarized their average methylation level as a
489 hypomethylation index.

490 We then searched for DMRs in transcriptionally active regions whose methylation levels were
491 associated with this hypomethylation index using the bumphunting method implemented in the R
492 package minfi [66]. In the NT1 $CD4^+$ T-cell analysis, the ratio of naïve $CD4^+$ T cells (naïve
493 CD4/CD4) estimated using the DNA Methylation Age Calculator, together with age, sex, and
494 disease status, was included as covariates. In the NT1 $CD8^+$ T-cell analysis, the ratio of naïve
495 $CD8^+$ T cells (naïve CD8/CD8), age, sex, and disease status were included as covariates. In the
496 MS analysis, estimated immune cell proportions (CD8, NK, B cell, monocyte, granulocyte, naïve

497 CD4, and naïve CD8) selected to minimize multicollinearity, along with age, sex, and disease
498 status, were included as covariates.
499 Only DMRs in which more than 70% of the CpGs were located within transcriptionally active
500 regions, as defined in this study, were retained for further analysis. To ensure the robustness of
501 the analysis, probes located within the *HLA* region (chr6:27,000,000–34,000,000, hg19) were
502 excluded. Genes linked to CpG sites within each DMR through eQTM relationships were
503 assigned to the corresponding DMRs based on previously reported eQTM datasets [41].
504

505 Estimation of T-cell receptor clonality

506 For NT1, T-cell receptor (TCR) clonality was estimated using whole-blood RNA-seq data
507 derived from the same samples used in the DNA methylation analysis. Details of the RNA-seq
508 data have been described previously [25]. TCR repertoires were reconstructed using TRUST4 [43].
509 Only productive TCR sequences were included in the analysis. Clonotypes were defined based
510 on CDR3 nucleotide sequences, following the approach described in the original TRUST4 study.
511 Clonality was calculated as 1 − normalized Shannon entropy of clonotype frequencies. Clonality
512 metrics were calculated separately for the TRA and TRB repertoires.
513

514 Prioritizing disease-associated genes through eQTM mapping

515 To prioritize disease-associated genes from disease-related hypomethylated loci, we utilized

516 eQTM relationships [41]. Detailed methods are described in the Supplementary Methods.

517

518 Bioinformatics workflow and command lines

519 The command lines used in this study are listed in Additional file 2. The Python programs are

520 also available on GitHub (https://github.com/mihshimada/Hypomethylation_immune).

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.

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.

The command lines used in this study are listed in Additional file 2. The python programs are also available on GitHub (https://github.com/mihshimada/Hypomethylation_immune).

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Competing interests

The authors declare that they have no financial arrangements or connections related to the content of this article to disclose.

Dr Makoto Honda has received consulting fees from Takeda Pharmaceutical Co. Ltd.; speaker honoraria from Aculys Pharma, Inc.; and editorial supervision honoraria from Alfresa Pharma Corporation, all unrelated to the submitted work.

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

556 MS and YO designed the study. MH collected NT1 samples. MS, TM, and YH carried out

557 experiments and statistical analyses. YH, TM, MH, TK, and KT interpreted the results. MS and

558 YO drafted the manuscript, and all authors contributed to the final version of the paper.

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. Shimada, M. *et al.* DNA methylation and multi-omics profiling of T cells uncovers chemotactic pathways and proliferation-linked hypomethylation in narcolepsy type 1. *Sleep* **48**(12) (2025).
26. 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).
27. 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).
28. 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).
29. Campagna, M.P. *et al.* Whole-blood methylation signatures are associated with and accurately classify multiple sclerosis disease severity. *Clin Epigenetics* **14**, 194 (2022).
30. 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).
31. 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).
32. Ulrich, A. *et al.* Blood DNA methylation profiling identifies cathepsin Z dysregulation in pulmonary arterial hypertension. *Nat Commun* **15**, 330 (2024).

630 33. Zhang, L. *et al.* Epigenome-wide meta-analysis of DNA methylation differences in prefrontal
631 cortex implicates the immune processes in Alzheimer's disease. *Nat Commun* **11**, 6114 (2020).
632 34. Schultz, M.D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation
633 variation. *Nature* **523**, 212-6 (2015).
634 35. Loyfer, N. *et al.* A DNA methylation atlas of normal human cell types. *Nature* **613**, 355-364
635 (2023).
636 36. Shimada, M., Miyagawa, T., Toyoda, H., Tokunaga, K. & Honda, M. Epigenome-wide
637 association study of DNA methylation in narcolepsy: an integrated genetic and epigenetic
638 approach. *Sleep* **41**(2018).
639 37. Rodríguez-Ubreva, J. *et al.* Inflammatory cytokines shape a changing DNA methylome in
640 monocytes mirroring disease activity in rheumatoid arthritis. *Ann Rheum Dis* **78**, 1505-1516
641 (2019).
642 38. Rhead, B. *et al.* Global expression and CpG methylation analysis of primary endothelial cells
643 before and after TNFa stimulation reveals gene modules enriched in inflammatory and
644 infectious diseases and associated DMRs. *PLoS One* **15**, e0230884 (2020).
645 39. Cardenas, H. *et al.* TGF-β induces global changes in DNA methylation during the epithelial-
646 to-mesenchymal transition in ovarian cancer cells. *Epigenetics* **9**, 1461-72 (2014).
647 40. Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the
648 Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *Am J Hum*
649 *Genet* **103**, 654-665 (2018).
650 41. Keshawarz, A. *et al.* Expression quantitative trait methylation analysis elucidates gene
651 regulatory effects of DNA methylation: the Framingham Heart Study. *Sci Rep* **13**, 12952
652 (2023).
653 42. Oliva, M. *et al.* DNA methylation QTL mapping across diverse human tissues provides
654 molecular links between genetic variation and complex traits. *Nat Genet* **55**, 112-122 (2023).
655 43. Song, L. *et al.* TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq
656 data. *Nat Methods* **18**, 627-630 (2021).
657 44. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and
658 functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**,
659 D638-D646 (2023).
660 45. Banchereau, R. *et al.* Personalized Immunomonitoring Uncovers Molecular Networks that
661 Stratify Lupus Patients. *Cell* **165**, 551-65 (2016).
662 46. Nakano, M. *et al.* Distinct transcriptome architectures underlying lupus establishment and
663 exacerbation. *Cell* **185**, 3375-3389.e21 (2022).
664 47. Kaech, S.M., Wherry, E.J. & Ahmed, R. Effector and memory T-cell differentiation:
665 implications for vaccine development. *Nat Rev Immunol* **2**, 251-62 (2002).

666 48. Hardy, R.R. & Hayakawa, K. B cell development pathways. *Annu Rev Immunol* **19**, 595-621
667 (2001).
668 49. Rådinger, M. *et al.* Local proliferation and mobilization of CCR3(+) CD34(+) eosinophil-
669 lineage-committed cells in the lung. *Immunology* **132**, 144-54 (2011).
670 50. Amin, K., Janson, C. & Bystrom, J. Role of Eosinophil Granulocytes in Allergic Airway
671 Inflammation Endotypes. *Scand J Immunol* **84**, 75-85 (2016).
672 51. Vilkaitis, G., Suetake, I., Klimasauskas, S. & Tajima, S. Processive methylation of
673 hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J Biol Chem* **280**, 64-72
674 (2005).
675 52. Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase
676 methylates DNA processively with high preference for hemimethylated target sites. *J Biol*
677 *Chem* **279**, 48350-9 (2004).
678 53. Wang, Q. *et al.* Imprecise DNMT1 activity coupled with neighbor-guided correction enables
679 robust yet flexible epigenetic inheritance. *Nat Genet* **52**, 828-839 (2020).
680 54. Nishiyama, A. & Nakanishi, M. Navigating the DNA methylation landscape of cancer. *Trends*
681 *Genet* **37**, 1012-1027 (2021).
682 55. Flynn, J., Azzam, R. & Reich, N. DNA binding discrimination of the murine DNA cytosine-C5
683 methyltransferase. *J Mol Biol* **279**, 101-16 (1998).
684 56. Bashtrykov, P., Ragozin, S. & Jeltsch, A. Mechanistic details of the DNA recognition by the
685 Dnmt1 DNA methyltransferase. *FEBS Lett* **586**, 1821-3 (2012).
686 57. Kieffer-Kwon, K.R. *et al.* Interactome maps of mouse gene regulatory domains reveal basic
687 principles of transcriptional regulation. *Cell* **155**, 1507-20 (2013).
688 58. Lio, C.J. *et al.* TET enzymes augment activation-induced deaminase (AID) expression via 5-
689 hydroxymethylcytosine modifications at the. *Sci Immunol* **4**(2019).
690 59. Nickles, D. *et al.* Blood RNA profiling in a large cohort of multiple sclerosis patients and healthy
691 controls. *Hum Mol Genet* **22**, 4194-205 (2013).
692 60. Ma, Q. *et al.* Specific hypomethylation programs underpin B cell activation in early multiple
693 sclerosis. *Proc Natl Acad Sci U S A* **118**(2021).
694 61. Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative
695 and comparative epigenomics. *PLoS One* **8**, e81148 (2013).
696 62. Decato, B.E. *et al.* Characterization of universal features of partially methylated domains across
697 tissues and species. *Epigenetics Chromatin* **13**, 39 (2020).
698 63. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature*
699 **489**, 57-74 (2012).
700 64. Battram, T. *et al.* The EWAS Catalog: a database of epigenome-wide association studies.
701 *Wellcome Open Res* **7**, 41 (2022).

702 65. Narita, T. *et al.* Acetylation of histone H2B marks active enhancers and predicts CBP/p300
703 target genes. *Nat Genet* **55**, 679-692 (2023).
704 66. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of
705 Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).

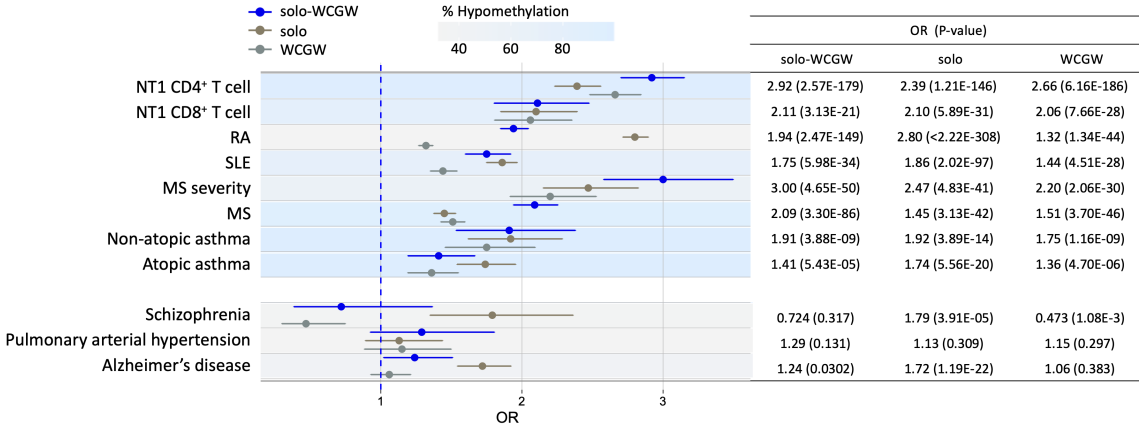
706 # Tables

707 Table 1. DMRs associated with hypomethylation index in NT1 $CD4^+$ T cells

DMR	Chr	Position (hg19)		Value	L	P-value	FWER	Replication		Genes in eQTM
		start	end					FWER<0.05	*P*<0.05	
DMR1	chr14	22966344	22975521	34.8	24	0	0	TRUE	TRUE	TRDV2:TRDC:TRAJ37:TRAV1-2
DMR2	chr15	98503878	98504173	-71.3	4	1.09E-05	0.068		TRUE	
DMR3	chr17	46656572	46660097	-33.9	22	2.26E-05	0.15		TRUE	HOXB2:HOXB3:HOXB4:HOXB-AS1
DMR4	chr6	42928079	42928920	-31.2	20	2.56E-05	0.17		TRUE	GNMT:PEX6:AL035587.1
DMR5	chr20	36148604	36149022	-28.5	20	2.59E-05	0.172		TRUE	

708 **Abbreviations:** DMR, differentially methylated region; FWER, family-wise error rate; eQTM, expression quantitative trait methylation; L, number
709 of CpGs within the DMR.

Figures



	OR (P-value)		
	solo-WCGW	solo	WCGW
NT1 CD4+ T cell	2.92 (2.57E-179)	2.39 (1.21E-146)	2.66 (6.16E-186)
NT1 CD8+ T cell	2.11 (3.13E-21)	2.10 (5.89E-31)	2.06 (7.66E-28)
RA	1.94 (2.47E-149)	2.80 (<2.22E-308)	1.32 (1.34E-44)
SLE	1.75 (5.98E-34)	1.86 (2.02E-97)	1.44 (4.51E-28)
MS severity	3.00 (4.65E-50)	2.47 (4.83E-41)	2.20 (2.06E-30)
MS	2.09 (3.30E-86)	1.45 (3.13E-42)	1.51 (3.70E-46)
Non-atopic asthma	1.91 (3.88E-09)	1.92 (3.89E-14)	1.75 (1.16E-09)
Atopic asthma	1.41 (5.43E-05)	1.74 (5.56E-20)	1.36 (4.70E-06)
Schizophrenia	0.724 (0.317)	1.79 (3.91E-05)	0.473 (1.08E-3)
Pulmonary arterial hypertension	1.29 (0.131)	1.13 (0.309)	1.15 (0.297)
Alzheimer's disease	1.24 (0.0302)	1.72 (1.19E-22)	1.06 (0.383)

Fig. 1. Hypomethylation and the accumulation of hypomethylated CpG sites in the solo-WCGW context in immune-mediated diseases. The proportion of hypomethylated disease-associated CpG sites was examined in immune-mediated and non–immune-mediated diseases. Enrichment of hypomethylated CpG sites in the solo context (i.e., lacking nearby CpGs) and in the WCGW context (CpGs flanked by A or T) was evaluated. In the forest plots, the background color for each disease represents the proportion of hypomethylated CpG sites among all disease-associated CpGs, with darker blue indicating a higher degree of hypomethylation.

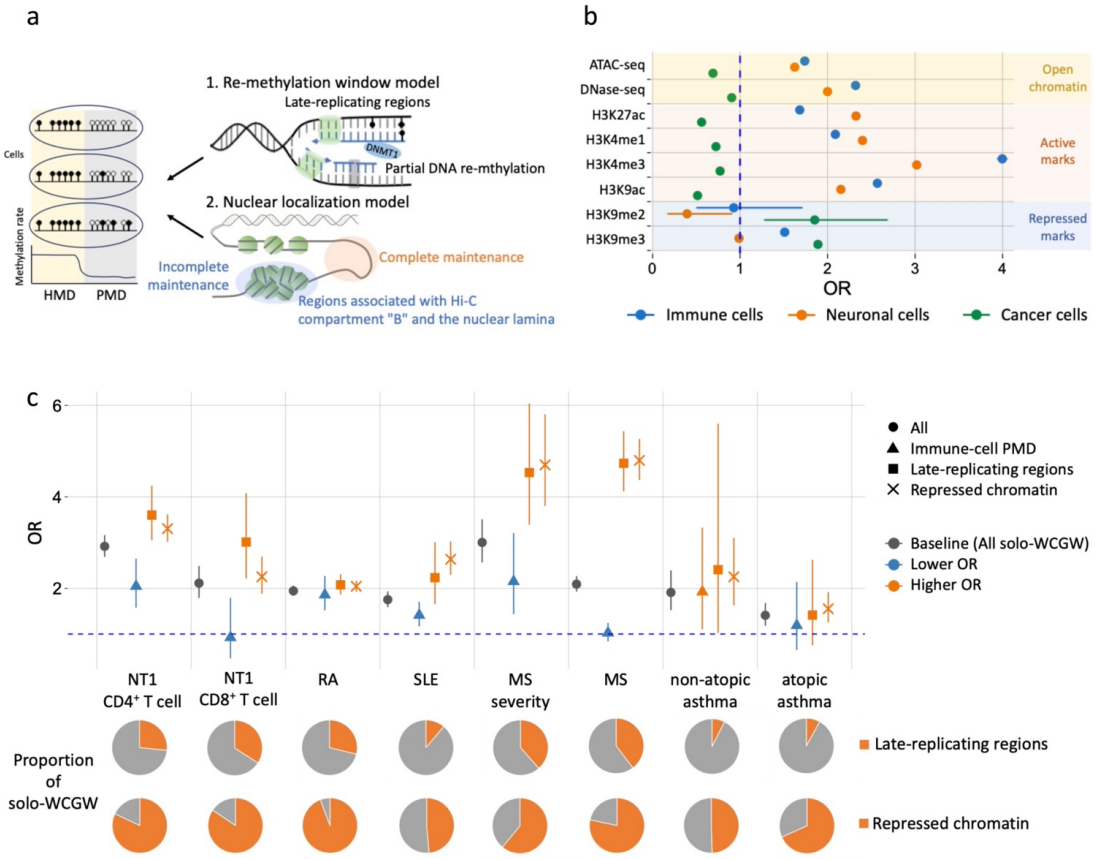


Fig. 2. Identification and characterization of regions in which passive demethylation of solo-WCGW is likely to occur. a. Schematic models proposed for passive demethylation of solo-WCGW sites. **b.** Overlap between the regions identified by ATAC-seq, DNase-seq, and histone ChIP-seq in immune cells and the PMDs in cancer, neural, and immune cells. **c.** 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.

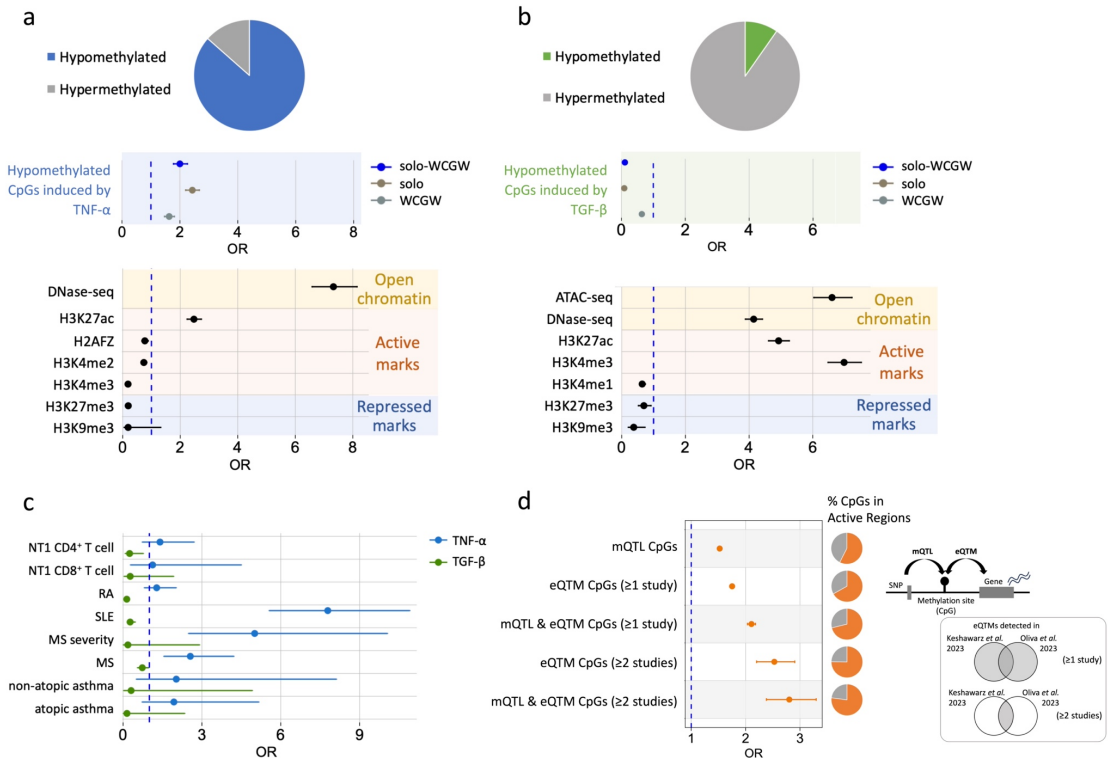


Fig. 3. Hypomethylation associated with cytokines and genetic variants is enriched in transcriptionally active regions. a. Changes in DNA methylation upon TNF-α treatment in human umbilical vein endothelial cells. The pie chart shows the proportion of CpG sites undergoing hypomethylation among those with altered methylation after TNF-α treatment. The middle forest plot shows the ORs for solo-WCGW among hypomethylated CpG sites. The bottom forest plot shows the ORs for enrichment of hypomethylated CpG sites in 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.** Odds ratios (ORs) for the overlap between disease-associated hypomethylated CpG sites and CpG sites hypomethylated in response to cytokine stimulation. **d.** ORs calculated by comparing the proportions of CpG sites with mQTLs and/or eQTMs located in transcriptionally active regions versus repressed chromatin. Pie charts show the proportion of CpGs located in mQTLs, eQTMs, or both that fall within active regions. For eQTMs, "≥1 study" indicates analyses using eQTMs detected in either of the two studies (upper Venn diagram), whereas "≥2 studies" indicates analyses using eQTMs detected in both studies (lower Venn diagram).

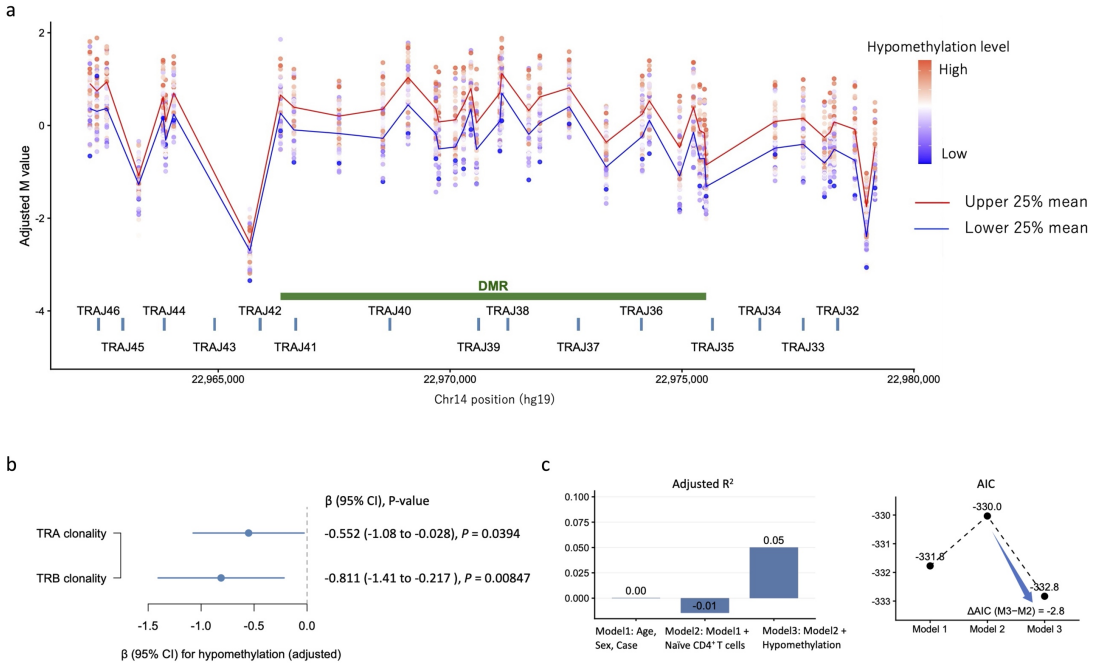


Fig. 4. Hypomethylation index correlates with methylation changes at the TRA locus and T-cell receptor clonality in NT1 CD4⁺ T cells. a. Methylation levels across the TRAJ locus on chromosome 14 in CD4⁺ T cells from NT1 samples. Points represent adjusted methylation values for individual CpG sites (covariates: age, sex, disease status, naïve CD4⁺ T-cell proportion), colored by hypomethylation level. Samples were stratified by the hypomethylation index, and lines indicate the mean methylation levels for the upper and lower quartiles. The green bar indicates the differentially methylated region (DMR) associated with the hypomethylation index. **b.** Association between the hypomethylation index and T-cell receptor clonality estimated from RNA-seq data. Effect sizes (β) with 95% confidence intervals are shown for TRA and TRB clonality. **c.** Comparison of regression models explaining T-cell clonality. Inclusion of the hypomethylation index improved model fit, as indicated by the increase in adjusted R^2 and the reduction in AIC.

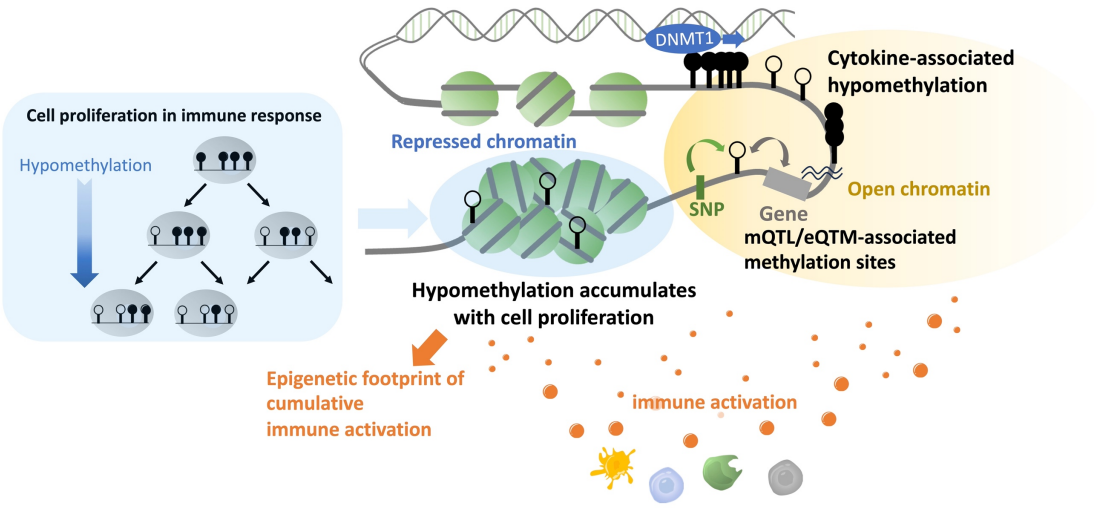


Fig. 5. Model of replication-associated hypomethylation as an epigenetic footprint of cumulative immune activation. Immune stimulation directly influences methylation in transcriptionally active regions, where regulatory changes such as cytokine signaling and genetic effects occur. In contrast, CpG sites in repressed chromatin are less directly affected by regulatory processes. Instead, immune activation promotes immune cell proliferation, leading to replication-associated passive demethylation that accumulates particularly at solo-WCGW sites. Because repressed chromatin is less exposed to other regulatory influences, this hypomethylation can accumulate as an epigenetic footprint of cumulative immune activation.