Multiomics analysis of narcolepsy T cells: global hypomethylation linked to T cell proliferation

**Mihoko Shimada[1, 2, 3, 6], Makoto Honda [3, 4], Yoshiko Honda [3], Tohru Kodama [3], Yuki Hitomi [5], Katsushi Tokunaga[1], *Taku Miyagawa [3,6]

[1] Genome Medical Science Project (Toyama), National Center for Global Health and Medicine (NCGM), Tokyo, Japan

[2] Center for Clinical Sciences, National Center for Global Health and Medicine (NCGM), Tokyo, Japan

[3] Sleep Disorders Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

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

[5] Department of Human Genetics, Research Institute, National Center for Global Health and Medicine (NCGM), Tokyo, Japan

[6] These authors are the paper's lead contacts

*Correspondence to: Taku Miyagawa, Ph.D.
Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, 2-1-6 Kamikitazawa, Setagaya, Tokyo 156-8506, Japan
Tel: +81-3-6834-2389
E-mail: miyagawa-tk@igakuken.or.jp

**Correspondence to: Mihoko Shimada, Ph.D.
Genome Medical Science Project (Toyama), National Center for Global Health and Medicine (NCGM), 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan
Tel: +81- 3-6205-6572
E-mail: mihshimada@ri.ncgm.go.jp

Summary

Narcolepsy type 1 (NT1) is a chronic sleep disorder caused by a loss of orexin-producing cells in the brain and involves autoimmune mechanisms, including the presence of autoreactive T cells. We performed a genome-wide DNA methylation analysis using $CD4^+$ and $CD8^+$ T cells of NT1 patients. Analysis of differentially methylated regions as well as multiomics analysis with genomic and transcriptomic data obtained from the same samples indicated that cell chemotaxis pathways are implicated as a cause in the pathogenesis of NT1. Additionally, we found global hypomethylation in both the T cells of NT1 cases ($CD4^+$: P = 1.69E-67; $CD8^+$: P = 4.83E-12). These NT1-associated hypomethylated sites were significantly more abundant in solo-WCGW (sequences without neighboring CpGs, where W is an A or T base; P = 9.87E-194). Solo-WCGW tends to lose DNA methylation over the course of cell divisions, suggesting enhanced T cell proliferation in NT1.

Keywords

narcolepsy, multiomics, $CD4^+$ T cell, $CD8^+$ T cell, hypomethylation

Introduction

Narcolepsy type 1 (NT1) is a chronic sleep disorder characterized by frequent lapses into sleep and cataplexy. Both genetic and environmental factors contribute to the development of NT1 [1-3]. Selective loss of orexin (hypocretin)-producing cells has been observed in the lateral hypothalamus of NT1 postmortem brains, indicating that the loss of these cells is involved in the onset of NT1 [4-6]; however, the mechanism has not yet been clarified. NT1 is strongly associated with human leukocyte antigen (*HLA*), and more than 90% of patients with NT1 carry the *HLA- DQB1*06:02* allele [7-10]. Genome-wide association studies (GWASs) of NT1 have also identified other genetic factors, such as *TNFSF4* [11], *CCR1/CCR3* [12], *TCRB* [13], *ZNF365* [13], *TCRA* [11,13,14], *CTSH* [11], *P2RY11/DNMT1* [15], the *IL10RB-IFNAR1* locus [13], and *CPT1B/CHKB* [16]; most of these factors are immune-related. Exposure to Pandemrix, an H1N1 ASO3-adjuvanted vaccine against a pandemic influenza, was found to be an environmental factor for the development of NT1 [17,18], and some immunological abnormalities have been reported to trigger the development of NT1 [19]. Recent studies have also reported the involvement of $CD4^+$ T cells and $CD8^+$ T cells in NT1. In a study using sensitive cellular screens, hypocretin-specific autoreactive $CD4^+$ T cells were found in all of the NT1 patients, and hypocretin-specific autoreactive $CD8^+$ T cells were found in some of the NT1 patients; in contrast, these cells were found in only a small number of the healthy controls [20]. Another study has reported that the frequency of autoreactive $CD8^+$ T cells specific for NT1-related peptides, which are presented primarily by *HLA-DQB1*06:02*, was lower in healthy individuals with the disease susceptibility allele *HLA-DQB1*06:02* than in patients with NT1 or *HLA-DQB1*06:02*-negative healthy individuals [21]. In addition, it was reported that a higher percentage of $CD4^+$ and $CD8^+$ T cells release cytokines in response to hypocretin in children with NT1

69 [22], and that variations in the memory subsets of $CD4^+$ and $CD8^+$ T cells are higher in
70 patients with NT1 [23]. These results suggest that an autoimmune basis involving $CD4^+$ and
71 $CD8^+$ T cells is associated with the pathogenesis of NT1.

72 DNA methylation is one of the major epigenetic mechanisms. The DNA methylation
73 status affects gene expression, and it has been investigated in many diseases, because it
74 is a relatively stable modification (although it can be changed by some factors) and is
75 more readily measurable than gene expression, which is susceptible to various factors and
76 prone to change [24,25]. Causal mutations in DNA methyltransferase 1 (*DNMT1*) have been
77 found in families with atypical narcolepsy [26], and it has been reported that patients with
78 *DNMT1* mutations show a genome-wide methylation abnormality [27]. In addition, a
79 GWAS of NT1 has reported a significant association between single-nucleotide
80 polymorphisms (SNPs) near *DNMT1* and NT1 [15]. We previously performed an
81 epigenome-wide association study (EWAS) of DNA methylation using peripheral blood
82 samples, and we found that methylation around *CCR3* (C-C motif chemokine receptor 3)
83 and genes related to hormone secretion and monocarboxylic acid metabolism pathways
84 were associated with NT1; in addition, we also found that NT1-related methylation sites
85 were significantly located outside of CpG islands, 95% of which were hypomethylated in
86 NT1 [28]. The array we used in that study was the Infinium HumanMethylation27 BeadChip
87 (27K array; Illumina, San Diego, CA, USA), which could only measure methylation in
88 about 27,000 locations. Although DNA methylation level is reportedly correlated with
89 methylation status in neighboring sites, the correlation is not so strong as the linkage
90 disequilibrium of SNPs [29], and the analysis of 27,000 sites was considered to be
91 insufficient for understanding the entire genome. Furthermore, the DNA methylation
92 patterns are known to vary widely among cells and tissues [30], and our previous study using

only peripheral blood cells might have been insufficient for detecting subtle changes in methylation that occur in individual immune cells.

In the present study, we performed EWASs of DNA methylation using $CD4^+$ and $CD8^+$ T cells isolated from whole blood samples to detect NT1-related differentially methylated sites, as these two types of immune cells have been suggested to play an important role in the pathogenesis of NT1. We also performed RNA sequencing (RNA-seq) using whole blood from the same subjects for methylation analysis to examine NT1-associated gene expression, at least that at the time of blood sampling. Genome-wide genotyping was also carried out, and we performed multi-omics analysis using genome, methylome, and transcriptome data collected from the same subjects. Although gene expression is the most phenotypic of these omics data, we focused our analysis on methylation with consideration of the potential of DNA methylation as a form of epigenetic memory [31]. Although it is difficult to capture the gene expression changes that trigger the development of NT1 in real-time since there is an inevitable time lag between disease onset and diagnosis (including sleep studies and blood sampling), DNA methylation may be a particularly useful tool for studying disease onset as the effects of such abnormal immune events may be retained as epigenetic memory.

Results

DNA methylation analysis and characteristics of the samples based on DNA methylation

We examined the DNA methylation profiles of $CD4^+$ and $CD8^+$ T cells using the EPIC methylation array. Our research design is shown in Fig. 1. After probe filtering and data normalization, 746,913 probes were selected for the following analysis (Fig. S1). We performed principal component analysis (PCA) analysis using normalized β values for all methylation sites, and confirmed that the $CD4^+$ T cell-derived and $CD8^+$ T cell-derived samples fell into widely different clusters, and that there was no significant bias in the distribution of the cases and controls within each cluster (Fig. S2A, B). We next examined whether the methylation rate of the master transcription factor binding site of T cells explained most of the variations in the PCA, as has been reported in previous studies [32]. Variation in PC1 was found to be well-explained by the methylation rate of the T-bet binding sites, and variation in PC2 was found to be well-explained by the methylation rate of the GATA3 binding sites (Fig. S2C, D). These results indicated that the variability in the present sample mainly reflected differences in the T cell subtypes. Although the PCA results showed no relationship between T cell subtype variation and case or control bias, we further examined in more detail the presence of T cell subtypes within each of the $CD4^+$ and $CD8^+$ T cell samples by estimating the composition of these subtypes in the discovery set. Estimation using the MeDeCom method revealed the lowest cross-validation error when assuming six subtypes for $CD4^+$ T cells and four subtypes for $CD8^+$ T cells (Fig. S3). We then compared the proportion of each estimated subtype between the NT1 cases and controls, and found that only the proportion of latent methylation component (LMC) 4 in the $CD4^+$ T cell analysis was significantly smaller in the NT1 cases than in the controls (P = 0.0032; Fig. S4A, B). Furthermore, we verified whether

134 the samples analyzed as $CD4^+$ and $CD8^+$ T cells were correctly estimated to be the
135 respective T cells using a reference-based estimation method. The results showed that in
136 most of the samples, the estimated T cell types ($CD4^+$ T cells or $CD8^+$ T cells) were largely
137 consistent with the sorted T cell types, but in three control samples of $CD8^+$ T cells, we
138 found that $CD8^+$ T cells were not estimated to be the major component since there was a
139 higher percentage of other leukocytes (Fig. S4C). These three control samples were
140 excluded from the subsequent association analyses, because they could not be properly
141 evaluated for $CD8^+$ T cell methylation. Similarly, estimation of the T cell proportions was
142 also performed for the replication samples. In five NT1 samples and four control samples,
143 we found that $CD8^+$ T cells were not correctly estimated to be the major component, and
144 these samples were thus excluded from the subsequent association analyses (Fig. S4C).
145 We also estimated the methylation age in the discovery set, and confirmed a strong
146 correlation between the DNA methylation age and chronological age ($r = 0.913$; Fig. S5A,
147 B, C). The biological age was not increased in the NT1 cases when compared to the
148 controls (Fig. S5C). The DNA methylation age and chronological age were also strongly
149 correlated in the replication samples ($r = 0.949$; Fig. S5D).
150
151 *Differentially methylated positions and regions of NT1*
152 We devised two strategies to identify methylation associated with NT1 to minimize
153 false positives. One was to identify differentially methylated regions (DMRs), which are
154 not just single-point associations like differentially methylated positions (DMPs), but
155 regions that have a chain of related methylation sites. The other was to investigate the
156 DMPs that have been indicated to be associated with NT1 by other omics data.

157 First, we examined DMPs to explore the overall trends of the associations. We found
158 DMPs that were significantly associated with NT1 after Bonferroni correction and were
159 replicated with $P < 0.05$ in three loci in the $CD4^+$ T cell analysis; they were all localized
160 to methylation sites within the *HLA* region (Fig. 2A), and no corresponding significant
161 DMPs were detected in $CD8^+$ T cell analysis (Fig. 2B). In our analysis, probes with poor
162 performance were excluded; however, the *HLA* region, being highly polymorphic and
163 challenging to analyze due to variations in probe affinity caused by the polymorphisms,
164 was analyzed separately, and the findings have been reported in a dedicated manner
165 elsewhere [33]. Despite the lack of substantial differences in the proportions of $CD4^+$ T cell
166 and $CD8^+$ T cell subtypes between the case and control groups, inflation of the observed
167 *P*-values was evident in the quantile-quantile (Q-Q) plots for the analyses involving both
168 $CD4^+$ and $CD8^+$ T cells. When we categorized differentially methylated sites into those
169 that were hypermethylated or hypomethylated in NT1, it was discernible that the inflation
170 was more pronounced at sites of hypomethylation (Fig. 2C, D).
171 Next, we focused our investigation on DMRs to identify plausible methylation sites
172 associated with NT1. We identified 15 replicated DMRs in the $CD4^+$ T cell analysis, and
173 five replicated DMRs in the $CD8^+$ T cell analysis (Table 1). Of the DMRs detected in the
174 $CD8^+$ T cell analysis, three were also identified in $CD4^+$ T cells. One of these three DMRs
175 was located in a region extending from upstream to the first exon of the C-C motif
176 chemokine ligand 5 (*CCL5*) gene on chromosome 17 (Table 1). Another was located in
177 an area near the cytoplasmic linker associated protein 2 (*CLASP2*) gene, which included
178 a methylation site that was found to be associated with C-C motif chemokine receptor 4
179 (*CCR4*) gene expression in $CD4^+$ T cells (Table 1).
180

Integrative analysis of methylation and gene expression

To reduce type 1 errors in the detection of DMPs associated with NT1 from genome-wide data, we examined DMPs that have been indicated to be associated with NT1 by gene expression (Fig. 1B). As it is well-known that the expression levels of numerous genes are related to the DNA methylation levels of nearby methylation sites, we first searched for methylation status and gene expression pairs that correlated with each other (Methylation-Expression pairs) from previous studies with larger samples sizes that examined both DNA methylation and gene expression in $CD4^+$ and $CD8^+$ T cells throughout the entire genome (Table S1) [34-38]. The results of the EWAS for DMPs indicated that the majority of NT1-related methylation sites are common to both $CD4^+$ and $CD8^+$ T cells, although the strength of the association with NT1 was stronger in $CD4^+$ T cells (Fig. 2); consequently, our analysis mainly focused on $CD4^+$ T cells. We analyzed the results for $CD4^+$ T cells from the three previous studies (Study 1, 2, and 3) listed in Table S1. In total, 3,191 and 78,881 Methylation-Expression pairs with a false discovery rate (FDR) < 0.05 were detected in Study 1 and 2, respectively. In Study 3, results for both naive $CD4^+$ T cells and $CD4^+$ T cells after stimulation were available, so we analyzed both, and detected 4,272 and 26,725 Methylation-Expression pairs, respectively (Table S1). These Methylation-Expression pairs were combined for a total of 110,713 Methylation-Expression pairs. Utilizing these pairs, we searched for combinations where both the methylation status and gene expression were associated with NT1, and the direction of this relation aligned with the direction of the relation in the Methylation-Expression pairs detected from the previous studies. We set the following criteria for each analysis: NT1-related methylation with $P < 0.05$ and a combined rank within the top 1% in the NT1 EWAS for DMP; and NT1-related gene expression with $P < 0.05$ and fold

205 change (FC) > 25% in the RNA-seq analysis. As a result, four Methylation-Expression
206 pairs were detected (Table S2). Among these, two pairs were related to natural killer (NK)
207 cell granule protein 7 (*NKG7*), and the two other pairs were related to programmed cell
208 death 1 (*PDCD1*) and junctional adhesion molecule 2 (*JAM2*). All of these genes were
209 genes that showed a negative correlation between their expression levels and the
210 methylation levels of NT1-associated methylation sites, and in the NT1 cases, the
211 expression levels of all of these genes increased and the methylation rate at the
212 methylation sites decreased.

213 Although we also conducted an analysis for $CD8^+$ T cells, there were no previous
214 studies with data were available on the detection of methylation using whole-genome
215 bisulfite sequencing or the EPIC array and the measurement of gene expression.
216 Therefore, we conducted the analysis of $CD8^+$ T cells using the results of two previous
217 studies (Study 4 and 5) that utilized the Infinium Human Methylation 450K BeadChip
218 (450K array; Illumina; Table S1), which preceded the EPIC array (850K array), making
219 it difficult to compare the $CD8^+$ T cell analysis results with the $CD4^+$ T cell analysis
220 results. When the analysis was conducted under conditions similar to those for the $CD4^+$
221 T cell analysis, no Methylation-Expression pair that met the criteria was detected in the
222 $CD8^+$ T cell analysis. We then relaxed the conditions for detecting correlated Methylation-
223 Expression pairs from FDR < 0.05 to a nominal significance level of $P < 0.05$, and
224 conducted the analysis. As a result, we identified three Methylation-Expression pairs, one
225 of which was related to *NKG7,* which was also detected in the analysis of $CD4^+$ T cells.
226 Similar to the $CD4^+$ T cell analysis, its expression level was increased in the $CD8^+$ T cells
227 in NT1 (Table S3).
228

229 *Integrative analysis of methylation and genotype data*

230 To integrate the genotype and methylation data, we used information of the
231 methylation quantitative trait locus (mQTL; Fig. 1C). Although tissue- and cell-specific
232 mQTLs were present, in the context of QTL analysis, it has been demonstrated that larger
233 sample sizes enhance the reliability of the results and enable the detection of a greater
234 number of QTLs [39]. Furthermore, prior research has reported that 84% and 80% of the
235 effects of mQTLs detected in $CD4^+$ and $CD8^+$ T cells, respectively, exhibit directions
236 concordant with those detected in whole blood. Consequently, we employed data
237 pertaining to mQTLs identified in studies focusing on whole blood with sample sizes
238 exceeding 1,000, as detailed in Table S4, for our integrative analysis. As shown in Table
239 S4, Study 6 and 7 reported 12,689,548 and 11,165,559 mQTLs, respectively. Among
240 these mQTLs, we searched for mQTLs related to NT1 that met the following conditions
241 for both the involved SNP and methylation: for the SNP, a $P < 1E\text{-}3$ in the GWAS of NT1
242 [12] and an odds ratio (OR) that was in the same direction and greater than the OR in the
243 GWAS in the samples of the present study (discovery and replication samples, NT1: N =
244 42 and control: N = 42); and for the methylation, a combined rank within the top 1% in
245 the discovery samples that was replicated. As a result, in $CD4^+$ T cells, four mQTLs from
246 Study 6 and two mQTLs from Study 7 were detected, with no mQTLs commonly detected
247 in both studies (Table S5). The four mQTLs detected in Study 6 were all pairs consisting
248 of two nearby SNPs and two methylation sites, which were found to negatively correlate
249 with the expression of S100 calcium binding protein A4 (*S100A4*; Table S5).

250 When a similar analysis was conducted for $CD8^+$ T cells, no mQTL showing
251 relationships between NT1 and both methylation and SNP was found. When the criteria
252 for SNPs related to NT1 in the GWAS were relaxed from $P < 1E\text{-}3$ to $P < 0.01$, two

253 mQTLs were detected (Table S6). However, in $CD8^+$ T cells, the correlation of SNPs and
254 methylation pairs was weaker than that in the $CD4^+$ T cells.
255
256 *Pathway analysis of genes affected by methylation sites associated with NT1*
257 Through the detection of DMRs and the integrative analysis of gene expression and
258 genotype data, various associated methylation regions or sites were identified, especially
259 in the analysis of $CD4^+$ T cells. Consequently, we conducted a pathway analysis of genes
260 associated with the methylation rates at these methylation sites by using the Methylation-
261 Expression information obtained from the $CD4^+$ T cell integrative analysis. The results
262 revealed the involvement of pathways related to immune responses, chemotaxis, and cell
263 migration (Fig. 2E and Table S7).
264
265 *Characterization of NT1-related hypomethylated sites*
266 In the process of conducting the association analyses for the identification of DMPs,
267 a significant inflation in *P*-values was observed across the analyses of both cell types,
268 with this phenomenon being more pronounced in $CD4^+$ T cells. Additionally, it was found
269 that this inflation predominantly occurred at methylation sites exhibiting
270 hypomethylation in NT1. Therefore, we explored the characteristics of these
271 hypomethylated sites in NT1 to determine why such a pronounced trend of
272 hypomethylation was observed in NT1.
273 In the examination of the prevalence of hypomethylation, the results indicated that for
274 the discovery set, across all thresholds ranging from a nominally significant $P = 0.05$ to
275 1E-5, more than 97% of the replicated NT1-associated methylation sites exhibited
276 hypomethylation in the $CD4^+$ T cells (Table 2). In contrast, 61% of all analyzed

methylation sites were hypomethylated in NT1. When comparing this ratio to the proportion of NT1-associated methylation sites that were hypomethylated in NT1, the results showed that the latter proportion was significantly larger (discovery set: probes with $P < 0.01$; $P = 1.69E-67$; Table 2). A similar trend was observed in $CD8^+$ T cells, in which over 80% of the NT1-associated methylation sites were hypomethylated in NT1 (Table 3). We then examined the regions in which NT1-associated hypomethylated sites were common with a $P < 0.01$ in the discovery set and were replicated with a $P < 0.05$. In both T cell analyses, we found that NT1-related hypomethylated sites were significantly less frequent within CpG islands ($CD4^+$ T cells: $P = 3.03E-115$ and $CD8^+$ T cells: $P = 9.30E-34$), and that they became more frequent with increasing distance from the CpG islands in the shores and shelves (Fig. 3A and Table S8). Furthermore, NT1-associated hypomethylated sites were particularly scarce in exon and promoter regions (exons, $CD4^+$ T cells: $P = 6.70E-26$ and $CD8^+$ T cells: $P = 7.98E-13$; Promoter_inactive, $CD4^+$ T cells: $P = 2.99E-22$ and $CD8^+$ T cells: $P = 1.61E-09$; Fig. #B and Table S9). When compared to the related regions from ChIP-seq in the T cells, NT1-related hypomethylated sites were significantly less frequent in regions marked by histone modifications associated with transcriptional activation, such as H3K4me3 and H3K27ac (Fig. #C and Table S10). Interestingly, despite the low overlap between NT-associated hypomethylated sites in the $CD4^+$ T cell analysis (N = 3,259) and those in the $CD8^+$ T cell analysis (N = 897), with only 38 sites in common, both shared the characteristic that they were less common in regions likely to influence gene expression, such as CpG islands, promoter regions, and open chromatin.

To investigate the relationships between NT1-associated hypomethylated sites, we generated heatmaps of the methylation rates for the hypomethylated sites in each sample.

301 The analysis revealed a consistent pattern: individuals with a higher degree of
302 hypomethylation showed this trait across almost all of these methylation sites, while those
303 with a lower degree of hypomethylation displayed higher methylation across the board
304 (Fig. S6). Consequently, the average methylation rate of these hypomethylated sites was
305 used as an indicator to represent the degree of hypomethylation in each sample. We
306 investigated the association between the degree of hypomethylation for each sample and
307 clinical information, including age, sex, and body mass index (BMI). Sex and BMI did
308 not show an association with the average value of hypomethylation in the discovery set
309 (Fig. S7A, B). In the replication set, the BMI appeared to be weakly correlated with the
310 average value of hypomethylation (R^2 = 0.396); this may be attributable to the fact that
311 the BMI was significantly higher in the NT1 cases than in the controls in the replication
312 set. Age was weakly correlated with the average value of hypomethylation, and a
313 tendency was observed for increased hypomethylation with higher age (R^2 = 0.123-0.467;
314 Fig. S7C). However, since age alone cannot explain the hypomethylation seen in NT1,
315 we next examined the relationship between hypomethylation and various indicators that
316 could be calculated with genome-wide methylation data using the DNA Methylation Age
317 Calculator [40] and MethylDetectR [41,42] (Table S11). The analysis was adjusted for age. It
318 should be noted that some of these indicators are based on values calculated assuming the
319 input data is from whole blood, but our methylation data was obtained from fractionated
320 T cells; thus, the reliability of these estimates was verified in the following section using
321 whole blood methylation analysis data that we previously obtained with the 27K array.
322 The analysis results showed that in the $CD4^+$ T cells, granzyme A (GZMA) demonstrated
323 the strongest association with hypomethylation (P = 2.26E-10), followed by correlations
324 with hypoSC (a methylation-based mitotic index; P = 6.08E-06; Fig. 3D, E and Table

S12). In $CD8^+$ T cells, hypoSC showed the strongest association with hypomethylation (P = 1.61E-12), followed by a correlation with GZMA (P = 3.44E-08; Fig. 3D, E and Table S13).

Investigation of NT1-related hypomethylation in whole blood

As described in the previous section, when exploring the indicators showing an association with hypomethylation in NT1 in the $CD4^+$ and $CD8^+$ T cells, some of the indicators were based on the assumption that the input data for estimation are methylation data derived from whole blood; thus, we also conducted studies using methylation data obtained from whole blood. We had previously reported that many NT1-related methylation sites were hypomethylated in NT1 when using whole blood for the analysis [28]. In the present study, to examine other details more thoroughly, we reanalyzed the data. As shown in Fig. S8A, the P-values of the methylation sites hypomethylated in NT1 were significantly inflated when compared to the expected values. In a manner similar to our analysis of T cells, we then examined the relationship between the average methylation rates of hypomethylated sites associated with NT1 and the indicators calculated using the DNA Methylation Age Calculator [40] and MethylDetectR [41,42] to confirm the reproducibility of the results in the T cells. In our analysis, we focused particularly on hypoSC and GZMA. As a result, no significant association with the hypomethylation rate was observed when only age was adjusted for (Tables S12, S13 and Fig. S8B, C). However, upon reviewing the plots and adjusting for the effect of NT1, nominally significant associations between hypomethylation in NT1 and both hypoSC and GZMA were observed (hypoSC: P = 1.14E-4 and GZMA: P = 1.56E-2; Table S14 and Fig. S8D, E).

Prevalence of solo-WCGW CpGs in NT1-associated hypomethylated sites

The hypoSC, an indicator related to cell mitosis, was among the indicators that showed an association with hypomethylation at NT1-related methylation sites. Previous studies have reported that DNA methylation at CpG sites is more likely to be lost in regions where there are fewer neighboring CpG sites, particularly in CpGs within the "WCGW" context (where W represents an A or T base; Solo-WCGW) [43]. Following the previous study, we investigated the relationship between NT1-related hypomethylated sites and the number of CpG within windows of ±35 bp as well as the sequence context, including the bases on either side of the CpG (Table S15). We found that in both types of T cells, NT1-associated hypomethylated CpG sites were significantly more prevalent in solo-WCGW ($CD4^+$ T cells: P = 9.87E-194, OR = 3.07 and $CD8^+$ T cells: P = 1.46E-22, OR = 2.20; Table 4). Furthermore, upon detailed examination, it was found that NT1-associated hypomethylated sites were particularly prevalent within CpG sites located in the sequence of ACGT (Fig. S9).

364 **Discussion**

365 In this study, we conducted an EWAS in $CD4^+$ and $CD8^+$ T cells from patients with
366 NT1, and found that in both types of T cells, methylation sites associated with chemokine-
367 related genes, such as *CCR4* and *CCL5*, were related to NT1. Particularly in $CD4^+$ T cells,
368 genes with expression levels that correlate with the methylation sites were predominantly
369 involved in pathways of immune responses and cell chemotaxis. Additionally, we
370 discovered the global hypomethylation of NT1-associated methylation sites in both cell
371 types, and that the degree of hypomethylation was well-correlated with the estimated
372 values of GZMA and hypoSC, indicators of cell mitosis. NT1-associated hypomethylated
373 sites were significantly more prevalent in solo-WCGW context, which is known to lose
374 methylation modifications easily during cell division, suggesting that cell division might
375 be enhanced in the T cells of NT1 patients.

376 When we searched for methylation sites associated with NT1, to minimize false
377 positives, we adopted two strategies. Firstly, instead of focusing only on single-point
378 associations, *i.e.,* DMPs, we focused on DMRs. Secondly, even for single-point DMP
379 associations, we sought to identify the DMPs that have been indicated to be associated
380 with NT1 by other omics data. In the analysis for DMR identification, more DMRs were
381 detected in the analysis of $CD4^+$ T cells than in the analysis of $CD8^+$ T cells (Table 1).
382 However, most DMRs were found to be commonly associated in both cell types when
383 including those that showed an association in either the discovery or replication analysis.
384 Among these, especially in the replicated DMRs of both cell types, there were DMRs
385 near *CCL5* and DMRs affecting the gene expression of *CCR4*. Both of these genes are
386 chemokine-related, and considering that the previous GWAS of NT1 reported
387 associations with a SNP, *i.e.,* rs3181077, located between other chemokine-related genes,

388 *i.e., CCR1* and *CCR3*, and considering that the observed decrease in their expression
389 levels in NT1, *CCL5* and *CCR4* represent highly intriguing candidate genes. *CCL5*, which
390 is also known as regulated on activation, normal T cell expressed and secreted (RANTES),
391 is a ligand for *CCR1*, *CCR3,* and *CCR5*, and it acts as a chemoattractant for monocytes,
392 memory T-helper cells, and eosinophils [44]. In this study, we were unable to find an
393 association between the upstream DMR of *CCL5* and the gene expression level of *CCL5*
394 itself, which was likely due to the very low expression levels of *CCL5* in the analyzed
395 data. An investigation into the expression in immune cells using DICE (the database of
396 immune cell expression, expression quantitative trait loci (eQTLs), and epigenomics) [45]
397 revealed that the average transcripts per million (TPM) was less than 1 in all cell types.
398 However, the single-cell analysis results showed expression primarily in $CD4^+$ T cells [45].
399 If we consider that the expression levels of genes with low expression levels, or in cases
400 where *CCL5* expression levels are increased by some kind of immune system activation,
401 there is a reasonable possibility of an association between gene expression and the NT1-
402 associated DMRs in the *CCL5* region identified in this study. *CCR4* is a chemokine
403 receptor for TARC/CCL17 and MDC/STCP-1/CCL22, which are mainly expressed by T
404 cells [46]. *CCR4* plays a crucial role in regulating immune balance, and is highly expressed
405 in activated regulatory T cells; thus, it is increasingly being recognized as a potential
406 target for cancer therapy [46,47].

407 The involvement of chemotaxis in NT1 was also suggested from the observed
408 association with *S100A4* in the results of the integrated analysis of mQTLs using genotype
409 data of $CD4^+$ T cells (Table S5). *S100A4* is a member of the Ca(2+)-dependent S100
410 protein family and a transcription factor that regulates cellular motility. It is known that
411 *S100A4* promotes filament depolymerization and chemotaxis through its direct

interaction with soluble myosin-IIA, leading to the formation of stable protrusions [48]. The gene expression of *S100A4* showed a negative correlation with methylation sites that were hypomethylated in the NT1 patient group, suggesting the possibility that, at least at the onset of NT1, the expression of *S100A4* was increased in the $CD4^+$ T cells of the patients with NT1. A common issue in normal DNA methylation analysis is the difficulty in determining whether the associated factors are causes or consequences of the disease. In our analysis, it was also challenging to distinguish whether the associations found with many methylation sites were causes or consequences of NT1; however, the results of the integrated analysis with genotype data indicated causal effects, because genetic factors can be considered as causal for the disease. Therefore, it is highly likely that changes in the methylation rates affected by genetic variants (mQTLs) contributed to the disease onset. Considering this in conjunction with previous reports suggesting associations with chemokines derived from GWAS SNP associations [12], chemokine-related pathways could be a significant cause of NT1 development. Further detailed examinations of the chemokine pathways are necessary.

In this study, we observed global hypomethylation at NT1-related methylation sites in both $CD4^+$ and $CD8^+$ T cells (Fig. 2 and Tables 2, 3). This trend was not only seen in the replication set, but was also confirmed in the results of the methylation analysis for NT1 using whole blood that we had conducted previously [49]. In our investigation of hypomethylated sites, we found that they were particularly scarce in CpG islands, and that they became more prevalent further from the islands. It was also evident that they were significantly less common in gene promoter regions and intragenic areas (Fig. 3). Although the methylation of CpG islands in the upstream promoter regions of genes greatly influences gene expression, much remains unknown about the functions of the

436 methylation sites in such outside regions. While exploring why such hypomethylation
437 occurs and what factors are involved, we discovered that the degree of hypomethylation
438 was highly correlated with the levels of GZMA estimated from methylation data, as well
439 as with the hypoSC cell division index (Fig. 3 and Tables S12, S13). We confirmed that
440 there was very little overlap between the methylation sites used to calculate the degree of
441 hypomethylation ($CD4^+$ T cells: N = 3,259 and $CD8^+$ T cells: N = 897) and those used to
442 estimate indices, such as GZMA and hypoSC (N = 3,629), with only five sites in $CD4^+$ T
443 cells and one site in $CD8^+$ T cells that overlapped. GZMA is one of the most abundant
444 proteases present in the cytoplasmic granules of cytotoxic T cells and NK cells, and it is
445 known for its crucial role in inducing the death of infected cells [50]. However, in recent
446 years, there has been debate regarding its role in cytotoxicity, and it has been shown that
447 extracellular GZMA is involved in the regulation of inflammation and inflammatory
448 diseases [50]. The potential of GZMA to exacerbate inflammation and cause excessive
449 immune responses, sepsis, tumor promotion, and rheumatoid arthritis has been suggested
450 [50]. In the present study, the association of NT1 with cytotoxic granule exocytosis and
451 inflammation was also suggested by the association of *NKG7* identified through the
452 integrated analysis of methylation and gene expression data (Table S2). NK cell granule
453 protein 7 is a factor that controls lymphocyte granule exocytosis and subsequent
454 inflammation in a wide range of diseases, and it has been reported that *NKG7*, which is
455 particularly expressed in $CD4^+$ and $CD8^+$ T cells, plays a crucial role in promoting
456 inflammation in infections [51]. Although determining whether the greater impact on
457 hypomethylation stems from events like inflammation, associated with GZMA, or from
458 the promotion of cell division, associated with hypoSC, is challenging due to the high
459 correlation between their estimated values ($CD4^+$ T cells: $R^2 = 0.528$ and $CD8^+$ T cells:

460 R^2 = 0.706), it is entirely plausible that T cell proliferation progressed due to the
461 accelerated inflammation [52]. The possibility of increased hypomethylation as a result of
462 increased cell division was supported by the observation that NT1-related
463 hypomethylated sites were especially prevalent in solo-WCGW motifs (Table 4).
464 Previous studies have shown that hypomethylation occurs in only about half of the
465 genome, especially in CG sequences not accompanied by surrounding CpGs and flanked
466 by A or T bases, such as in WCGW motifs [43]. It is also known that hypomethylation is
467 more likely to occur in areas with low gene density, in areas with low GC density, and in
468 so-called compartment B regions [43,53]. These genomic characteristics overlap well with
469 the features of regions where NT1-related hypomethylated sites were predominantly
470 found. In other autoimmune diseases, particularly in systemic lupus erythematosus (SLE)
471 and multiple sclerosis, global hypomethylation in immune cells has been reported [54,55]. In
472 the case of SLE, there have been reports suggesting that hypomethylated sites are often
473 found upstream of immune-related genes, indicating that the mechanism may differ
474 slightly from that of the hypomethylation we observed in NT1 [54]. In multiple sclerosis,
475 the global hypomethylation sites were particularly numerous within promoter regions in
476 B cells, and they were reported to influence the expression of genes involved in the
477 activation of B cells, which in turn further activates B cells [55]. In contrast, the
478 hypomethylation observed in NT1 was thought to possibly result from the progression of
479 T cell division. Therefore, even though global hypomethylation is observed across
480 different immune-related disorders, the mechanisms involved appear to be different.

481 There are several limitations to this study. First, we obtained gene expression data
482 from the same samples from which we acquired methylation data and conducted an
483 integrated analysis, but the expression data was derived from the mRNA in whole blood.

If we had more samples available, it would have been preferable to obtain gene expression data from separated $CD4^+$ and $CD8^+$ T cells. Thus, we might have missed associations for some genes, especially those with varying expression patterns in different immune cells. Second, among the indicators showing a relationship with global hypomethylation in NT1, the estimated values for proteins, including GZMA, were based on predictions intended for methylation data derived from whole blood, but our methylation data was obtained from isolated T cells, leaving questions about the accuracy of these estimates. Although we have validated the factors of interest using methylation data from whole blood, this data was obtained using the 27K array, which has fewer probes. Therefore, it is desirable to confirm the reproducibility in the future using EPIC array data derived from whole blood. Thirdly, although we conducted a comprehensive methylation analysis using arrays, we cannot deny the possibility that methylation sites not covered by the arrays may be associated with NT1. In the future, when whole-genome bisulfite sequencing with a deeper depth than the sequencing depths possible now becomes available, more associated factors may be identified. Lastly, in this study, we first isolated $CD8^+$ T cells from the buffy coat using magnetic beads, and then separated $CD4^+$ T cells from the remaining blood components, resulting in the possibility that the $CD8^+$ T cells included $CD4^+CD8^+$ double-positive T cells. Furthermore, we did not examine the different subtypes of $CD4^+$ and $CD8^+$ T cells. The global hypomethylation associated with NT1 showed a high correlation with indicators, such as GZMA and hypoSC, suggesting that inflammation and increased cell division could explain much of the hypomethylation. However, there were still differences in the degree of hypomethylation between the case and control groups that could not be fully explained. Namely, when comparing the same hypoSC and GZMA values, the NT1 cases tended to show more hypomethylation than

the controls, particularly in the results of the $CD8^+$ T cells and in the replication analysis with the 27K array using whole blood data. These data suggest the possibility of hypomethylation in NT1 that cannot be explained by inflammation or increased cell division alone. Particularly, the significant fact that hypomethylation cannot be explained by GZMA and hypoSC alone in the whole blood data indicates that the presence of different subtypes of immune cells in NT1 compared to control individuals may be a contributing factor. Future single-cell analyses in NT1 could potentially reveal the details of the immune abnormalities that contribute to the onset of the disease.

In conclusion, the results of our DNA methylation analysis of $CD4^+$ and $CD8^+$ T cells in NT1 suggest a high likelihood that abnormalities in chemokine-related pathways are a major cause of disease onset. Furthermore, our results indicate that increased T cell division in NT1, possibly driven by heightened inflammatory responses, may lead to hypomethylation in regions of DNA where methylation is more easily disrupted. Future investigations, including single-cell analyses and single-cell DNA methylation analyses, are expected to reveal more details of the mechanism of NT1 onset.

523 **Methods**

524 *Subjects in the discovery and replication sets*

525 Our research design is shown in Fig. 1. Samples from 42 (discovery: 28, replication:
526 14) patients with NT1 and 42 (discovery: 28, replication: 14) control subjects without
527 significant sleep-disturbing events who underwent diagnostic sleep studies
528 (polysomnogram and the multiple sleep latency test) at Seiwa Hospital (Shinjuku, Tokyo,
529 Japan) or Koishikawa Tokyo Hospital (Bunkyo, Tokyo, Japan) were included. All subjects
530 were Japanese. The demographic characteristics are shown in Table S16 and S17. NT1
531 was diagnosed by sleep specialists according to the International Classification of Sleep
532 Disorders, 3rd edition [56], and all patients carried the *HLA-DQB1*06:02* allele. Fasting
533 blood was drawn at 7:00 am after a polysomnogram, and all subjects were drug-naïve or
534 had stopped medication that could affect sleep at least 1 week before the time of blood
535 collection. We used two EDTA-2Na-coated blood collection tubes: one for buffy coat
536 DNA extraction and the other for lymphocyte separation by density-gradient
537 centrifugation using Ficoll-Paque Plus (Merck Millipore, USA). The lymphocyte fraction
538 was washed twice with phosphate-buffered saline, resuspended in Cellbanker 2 (Zenogen
539 Pharma, Japan), and stored at -80°C until use. We simultaneously collected blood into
540 two Paxgene RNA collection tubes (Beckton-Dickinson, Franklin Lakes, NJ, USA), and
541 extracted RNA using a Paxgene Blood RNA kit (QIAGEN, Hilden, Germany) according
542 to the manufacturers' protocols. The extracted RNA was used for RNA-seq as described
543 below. This study received ethical approval from all relevant agencies, and written
544 informed consent was obtained from all participants.

545

546 *Isolation of $CD4^+$ and $CD8^+$ T cells and DNA extraction*

547 $CD4^+$ and $CD8^+$ T cells were extracted from the lymphocyte fraction using the
548 Dynabeads CD4 Positive Isolation Kit and Dynabeads CD8 Positive Isolation Kit
549 (Invitrogen, Waltham, MA, USA) according to the manufacturer's protocols. First, $CD8^+$
550 T cells were isolated from the buffy coat, and $CD4^+$ T cells were isolated from the
551 remaining blood cell components. DNA was extracted from both types of T cells using
552 the QIAamp DNA Mini Kit (QIAGEN).
553
554 *DNA methylation analysis*
555 DNA was first bisulfite-converted, then the genome-wide methylation profile was
556 examined with a DNA methylation array (EPIC array; Infinium® Methylation EPIC
557 BeadChip, Illumina). The methylation arrays were imaged using a high-precision scanner
558 (iScan System, Illumina); the scan data (idat files) were captured with the ChAMP R
559 package [57-59]. The methylation rate was calculated as the β value and the M values (the
560 logit transformations of the β values) were used in the case-control analysis. Detection P
561 values were calculated using a method reported by Heiss *et al.* [60].
562
563 *Filtering and normalization of the methylation data*
564 Poor-performing probes were excluded from the analysis according to a method
565 reported by Zhou *et al.* (Fig. S1, S10) [61]. Data normalization involved three steps: (1)
566 correcting the bias between Infinium I and II assay systems using beta mixture quantile
567 dilation [62], (2) applying quantile normalization to adjust for variations in methylation rate
568 distributions between samples, and (3) using ComBat [63] with T cell type ($CD4^+$ or $CD8^+$)
569 as a covariate to correct for batch effects.
570

571 *Quality control of the methylation data*

572 The following three analyses were performed for quality control:

573 1) PCA using methylation data for all sites to check for bias in the distribution of the
574 PCA plot between case and control samples. In a previous study, when PCA was
575 performed using genome-wide methylation rate data for $CD4^+$ T cells, the principal
576 components were reported to be mainly explained by the methylation rate of the GATA
577 binding protein 3 (GATA3) binding sites [32]. GATA3 is the master transcription factor of
578 T cells, and it controls the differentiation and function of each T cell subset [64]. Therefore,
579 we additionally checked whether the methylation rate of the binding sites of master
580 transcription factors, *i.e.,* T-bet and GATA3, explained the main component in our PCA
581 as well. We utilized previously reported ChIP-Seq data for T-bet and GATA3 [65], and for
582 the methylation rate of each master transcription factors' binding site, dimensionality
583 reduction by PCA was performed and PC1 was used as a feature.

584 2) Estimation of the cellular composition within each sample. In this study, although
585 we separated $CD4^+$ and $CD8^+$ T cells for the methylation analysis, there are subsets of
586 each cell type that can be further subdivided. To confirm whether the composition of such
587 subsets differed between the case and control groups, we estimated the cellular
588 composition using MeDeCom, which does not require any reference methylation data [66].
589 In addition, we also performed reference-based cellular composition estimation as
590 reported by Horvath [40] to estimate the percentage of $CD4^+$ T cells and $CD8^+$ T cells in the
591 samples to confirm that each sample was properly estimated to contain mainly $CD4^+$ T
592 cells or $CD8^+$ T cells. Samples in which the most abundant cell type was not correctly
593 estimated to be the expected $CD4^+$ T cells or $CD8^+$ T cells were excluded from the
594 subsequent association analyses.

595 3) Estimation of age using methylation data. The biological age can be estimated from
596 methylation data [40]. We examined whether the biological age was accelerated compared
597 to the chronological age in NT1 patients, and we also examined whether the ages of the
598 controls were properly estimated.
599
600 *Association analysis of the methylation data*
601 To detect NT1-associated methylation sites, we searched for DMPs and DMRs using
602 the *t* test and bump hunting method [67], respectively. For DMPs, since not only the
603 statistical significance from P values, but also a large difference in the methylation rate is
604 important for biological significance, we calculated a combined rank considering both
605 indices (the *P* value and $\Delta\beta$, which is the average β for the cases – average β for the
606 controls), and we considered the top-ranked NT1-associated methylation sites to be
607 particularly important. Association analyses and visualization of the data, such as Q-Q
608 plots, were conducted using R software (ver. 3.6.2).
609
610 *RNA sequencing*
611 We extracted RNA from the whole blood samples of the 42 patients with NT1 and 42
612 control subjects included in the discovery and replication sets for DNA methylation
613 analysis and performed RNA-seq. Sequencing libraries were prepared using TruSeq
614 Stranded Total RNA Ribo Zero (Illumina), and paired-end RNA-seq was performed on a
615 NovaSeq 6000 (Illumina). The FASTQ read data were trimmed using a quality score of
616 0.05 and a minimum length of 15 nt in CLC Genomics Workbench 21 (QIAGEN). Read
617 mapping was performed with the following parameters: mismatch cost, 2; insertion cost,
618 3; deletion cost, 3; length fraction, 0.8; similarity fraction, 0.8; and maximum number of

619 hits for a read, 10. The read count data were corrected by the trimmed mean of M values
620 (TMM) method implemented in the edgeR package [68-70]. Association analysis of the
621 expression data was performed using the likelihood ratio test with age, sex, and BMI as
622 covariates in the design matrix.
623
624 *Genotyping*
625 DNA samples that were included in the DNA methylation analysis and RNA-seq (NT1:
626 N = 42 and control: N = 42) were genotyped with Axiom Japonica array v2 (Toshiba,
627 Tokyo, Japan). Imputation analysis was performed using BEAGLE5.1 [71] with reference
628 panels from the 1000 Genomes Project (phase 3) and from the Tohoku Medical Megabank
629 14KJPN [72]. SNP filtering was performed with the following criteria: information score ≥
630 0.5; minor allele frequency ≥ 0.01; SNP call rate ≥ 95%; and Hardy-Weinberg equilibrium
631 P value > 0.01. Genotype data from a previous GWAS of NT1 were also used in the
632 analysis. The details of the GWAS data are described elsewhere [12,73].
633
634 *Integrative analysis of methylation and gene expression*
635 To perform an integrative analysis of the methylation and gene expression data, we first
636 detected pairs of methylation rate and gene expression that correlated with each other
637 (Methylation-Expression pairs) using data from previous studies with larger sample sizes
638 (Table S1) [34-38]. Next, we searched for pairs in which the correlated methylation rate and
639 gene expression were both associated with NT1 in the samples of this study, and in which
640 the direction of the association was consistent with the Methylation-Expression pair (Fig.
641 1B). We searched for previous studies that measured genome-wide methylation rates and
642 gene expression in the same sample of $CD4^+$ T cells or $CD8^+$ T cells, and had a sample

size of N > 50. For $CD4^+$ T cells, we included the three previous studies shown in Table S1, while for $CD8^+$ T cells, there were no studies that reported EPIC (Illumina) or sequence-based methylation data, and only data from the 450K array (Illumina) were available. Since our primary focus was on $CD4^+$ T cell analysis from the results of the association analysis of NT1, for the $CD8^+$ T cells, we conducted an integrative analysis with the 450K data as a preliminary analysis. For the methylation data of the previous studies, we used filtered and normalized data, if available. For studies with only raw methylation data, standard filtering and normalization of ChAMP were performed. For expression data, we used the data processed in each study, and gene identifiers on each platform were converted to Ensembl identifiers using information from GENCODE [74] and BioMart [75].

Correlated Methylation-Expression pairs were explored using Matrix eQTL [76] with age and sex as covariates. In the analysis of Study 1, age was estimated from the genome-wide methylation data and used as a covariate, because no age information was available for Study 1 (Table S1). In Study 3 (Table S1), only sex was used as a covariate, because the subjects were children aged 0 to 2 years. We specifically focused on Methylation-Expression pairs in *cis* (<1 Mb), and identified pairs under two conditions: $P < 0.05$ and FDR < 0.05.

Among the detected Methylation-Expression pairs, we searched for pairs in which both the methylation rate and gene expression were related to NT1 (Fig. 1B). An NT1-associated methylation site, *i.e.,* a DMP, was defined as a site with a $P < 0.05$ and a combined rank within the top 1% among the discovery samples, and was replicated with a $P < 0.05$ in the replication samples. NT1-associated gene expression was defined as expression with a $P < 0.05$ and FC >25% using our RNA-seq data.

Integrative analysis of methylation and genotype data

To integrate the genotype and methylation data, we used information from the mQTL. An mQTL is where the methylation rate correlates with the genotype of the SNP, and it has been reported that the mQTL profiles are nearly identical between whole blood and each immune cell [77]. We searched for previous studies that used whole blood with a large sample size, and used the mQTL information reported in the two studies shown in Table S4 [77,78]. Among the reported mQTLs, we searched for those that had a methylation rate and SNPs that were both associated with NT1, and whose direction of association was consistent with the previous studies. For SNPs, we focused on those that were associated with NT1 in our previous GWAS to exclude SNPs that showed associations only in this study due to the small sample size (NT1: N = 42 and control: N = 42), and not in the previous GWAS of NT1 with larger sample size [12] (Fig. 1C). The detailed conditions for the mQTL search were as follows: an NT1-associated methylation site, *i.e.,* a DMP, was defined as a site with a $P < 0.05$ and a combined rank within the top 1% among the discovery samples, and was replicated with a $P < 0.05$ in the replication samples. NT1-associated SNPs were defined as those with a $P <$ 1E-3 in the GWAS of NT1, and in the present sample, when the OR > 1, those with an OR greater than the OR in the GWAS, and when the OR < 1, those with an OR smaller than the OR in the GWAS (Fig. 1C).

Pathway analysis

In the $CD4^+$ T cell analyses, since NT1-associated methylation sites were suggested in the analysis for DMRs and in the integrative analyses with gene expression data and genotype data, pathway analysis was conducted using genes whose expression was

691 affected by the methylation sites. The gene expressions affected by the methylation sites
692 were linked based on the relationship of Methylation-Expression pairs, as determined
693 through the integrative analysis of gene expression data and methylation data. Pathway
694 analysis was performed using MetaCore™ software (version 6.24 build 67895, Clarivate,
695 Boston, MA, USA), and the PANTHER Classification System (PANTHER 17.0) [79] was
696 used to validate the results from the MetaCore software. We then searched for significant
697 gene sets that met the following conditions in the Gene Ontology database [80,81]: i) FDR <
698 0.05 in the MetaCore analysis, and replicated in PANTHER with $P < 0.05$; ii) The
699 pathway includes three or more NT1-related genes; iii) The total number of objects in the
700 pathway is less than 3,000. The detected pathways were clustered using the k-means
701 method based on the presence or absence of NT1-related genes associated with each
702 pathway. The number of clusters, *k*, was set to four based on manual curation and intuitive
703 grouping observed within the detected pathways.
704
705 *Investigation of NT1-related hypomethylated sites*
706 Since a global hypomethylation trend was observed in NT1, we examined the NT1-
707 associated hypomethylation sites with a $P < 0.01$ in the discovery set that were replicated
708 with a $P < 0.05$. To explore the kinds of genomic regions where the hypomethylated sites
709 are commonly present, Ensembl Human Regulatory Features (GRCh38.p14) was used to
710 examine the positional relationship between hypomethylated sites and genomic
711 regulatory features, such as enhancers and promoters [82]. We extracted the data originating
712 from T cells (Table S18). GENCODE (version 28) data were used to determine the
713 positional relationship of regions related to genes [74]. We also used the functional genomics
714 data obtained from the ENCODE project [83] to investigate genomic features by Histone

ChIP-seq. The experimental matrix was filtered by the following conditions: status, released; perturbation, not perturbed; organism, *Homo sapiens*; cell, T cell; available file types: bed narrowPeak; and all audit category filters were applied. The positional information for CpG islands was based on Illumina's annotation file for the EPIC array. Since the genomic regulatory features from Gencode and Encode projects were based on hg38, the positional information of hypomethylated sites (hg19) was converted by the NCBI Genome Remapping Service for comparisons [84].

To visualize the relationships between hypomethylated sites, we generated heat maps, and confirmed that the methylation rates were generally low in samples with low levels of methylation in the NT1-associated hypomethylation sites. Since some subjects showed a generally high methylation status and others showed a generally low hypomethylation status across the selected DMPs, the average methylation rates of NT1-associated hypomethylation sites were calculated for each sample, serving as an indicator of the degree of hypomethylation. To identify factors that could explain this degree of hypomethylation, correlations with age, sex, and BMI were examined.

To examine factors associated with the degree of hypomethylation, we calculated various indicators that could be determined from genome-wide methylation data using the DNA Methylation Age Calculator [40] and MethylDetectR [41,42] (Table S11). We then conducted a linear regression analysis using age as a covariate to explore indices indicating an association with the degree of hypomethylation. Some indices were designed assuming that the methylation data were derived from whole blood as the input. Therefore, for the validation of the results, similar estimations were also conducted using data from an EWAS of whole blood from patients with NT1 that we had previously performed [28]. The prior study included 26 cases of NT1 and 20 controls, and the

methylation levels were measured using the 27K array (Illumina); the details of the analysis are described elsewhere [28].

Acknowledgements

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). 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.

Authors' contributions

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

Disclosures and competing interests

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

Data availability

The genome-wide methylation dataset supporting the conclusions of this article 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.

767 **References**

768 1 Khatami, R. *et al.* Monozygotic twins concordant for narcolepsy-cataplexy without any
769 detectable abnormality in the hypocretin (orexin) pathway. *Lancet* **363**, 1199-1200 (2004).
770 https://doi.org:10.1016/S0140-6736(04)15951-5
771 2 Mignot, E. Narcolepsy: Genetics, Immunology, and Pathophysiology. In: Kryger M, Roth
772 T, Dement W, editors. Principles and Practice of Sleep Medicine. 6th ed. p. 855-872
773 (Elsevier, 2017).
774 3 Mignot, E. Genetic and familial aspects of narcolepsy. *Neurology* **50**, S16-22 (1998).
775 4 Peyron, C. *et al.* A mutation in a case of early onset narcolepsy and a generalized absence
776 of hypocretin peptides in human narcoleptic brains. *Nat. Med.* **6**, 991-997 (2000).
777 https://doi.org:10.1038/79690
778 5 Thannickal, T. C. *et al.* Reduced number of hypocretin neurons in human narcolepsy.
779 *Neuron* **27**, 469-474 (2000).
780 6 Honda, M. *et al.* Absence of ubiquitinated inclusions in hypocretin neurons of patients
781 with narcolepsy. *Neurology* **73**, 511-517 (2009).
782 https://doi.org:10.1212/WNL.0b013e3181b2a6af
783 7 Mignot, E. *et al.* Complex HLA-DR and -DQ interactions confer risk of narcolepsy-
784 cataplexy in three ethnic groups. *Am J Hum Genet* **68**, 686-699 (2001).
785 8 Han, F. *et al.* HLA-DQ association and allele competition in Chinese narcolepsy. *Tissue*
786 *Antigens* **80**, 328-335 (2012). https://doi.org:10.1111/j.1399-0039.2012.01948.x
787 9 Hong, S. C. *et al.* DQB1*0301 and DQB1*0601 modulate narcolepsy susceptibility in
788 Koreans. *Hum Immunol* **68**, 59-68 (2007). https://doi.org:10.1016/j.humimm.2006.10.006
789 10 Roh, E. Y. *et al.* Association of HLA-DR and -DQ genes with narcolepsy in Koreans:
790 comparison with two control groups, randomly selected subjects and DRB1*1501-
791 DQB1*0602--positive subjects. *Hum Immunol* **67**, 749-755 (2006).
792 https://doi.org:10.1016/j.humimm.2006.03.024
793 11 Faraco, J. *et al.* ImmunoChip study implicates antigen presentation to T cells in
794 narcolepsy. *PLoS Genet* **9**, e1003270 (2013).
795 https://doi.org:10.1371/journal.pgen.1003270
796 12 Toyoda, H. *et al.* A polymorphism in CCR1/CCR3 is associated with narcolepsy. *Brain*
797 *Behav Immun* **49**, 148-155 (2015). https://doi.org:10.1016/j.bbi.2015.05.003
798 13 Han, F. *et al.* Genome wide analysis of narcolepsy in China implicates novel immune loci
799 and reveals changes in association prior to versus after the 2009 H1N1 influenza
800 pandemic. *PLoS Genet* **9**, e1003880 (2013). https://doi.org:10.1371/journal.pgen.1003880
801 14 Hallmayer, J. *et al.* Narcolepsy is strongly associated with the T-cell receptor alpha locus.
802 *Nat Genet* **41**, 708-711 (2009). https://doi.org:10.1038/ng.372

15 Kornum, B. R. *et al.* Common variants in P2RY11 are associated with narcolepsy. *Nat Genet* **43**, 66-71 (2011). https://doi.org:10.1038/ng.734

16 Miyagawa, T. *et al.* Variant between CPT1B and CHKB associated with susceptibility to narcolepsy. *Nat Genet* **40**, 1324-1328 (2008). https://doi.org:10.1038/ng.231

17 Nohynek, H. *et al.* AS03 adjuvanted AH1N1 vaccine associated with an abrupt increase in the incidence of childhood narcolepsy in Finland. *PLoS One* **7**, e33536 (2012). https://doi.org:10.1371/journal.pone.0033536

18 Dauvilliers, Y. *et al.* Increased risk of narcolepsy in children and adults after pandemic H1N1 vaccination in France. *Brain* **136**, 2486-2496 (2013). https://doi.org:10.1093/brain/awt187

19 Miyagawa, T. & Tokunaga, K. Genetics of narcolepsy. *Hum Genome Var* **6**, 4 (2019). https://doi.org:10.1038/s41439-018-0033-7

20 Latorre, D. *et al.* T cells in patients with narcolepsy target self-antigens of hypocretin neurons. *Nature* **562**, 63-68 (2018). https://doi.org:10.1038/s41586-018-0540-1

21 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). https://doi.org:10.1038/s41467-019-08774-1

22 Cogswell, A. C. *et al.* Children with Narcolepsy type 1 have increased T-cell responses to orexins. *Ann Clin Transl Neurol* **6**, 2566-2572 (2019). https://doi.org:10.1002/acn3.50908

23 Moresco, M. *et al.* Flow cytometry T cell profiling in a recent-onset narcoleptic type 1 child: a case report. *Sleep Med* **68**, 21-23 (2019). https://doi.org:10.1016/j.sleep.2019.08.017

24 Flanagan, J. M. Epigenome-wide association studies (EWAS): past, present, and future. *Methods Mol Biol* **1238**, 51-63 (2015). https://doi.org:10.1007/978-1-4939-1804-1_3

25 Berdasco, M. & Esteller, M. Clinical epigenetics: seizing opportunities for translation. *Nat Rev Genet* **20**, 109-127 (2019). https://doi.org:10.1038/s41576-018-0074-2

26 Winkelmann, J. *et al.* Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum Mol Genet* **21**, 2205-2210 (2012). https://doi.org:10.1093/hmg/dds035

27 Kernohan, K. D. *et al.* Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. *Clin Epigenetics* **8**, 91 (2016). https://doi.org:10.1186/s13148-016-0254-x

28 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). https://doi.org:10.1093/sleep/zsy019

29 Liu, Y. *et al.* GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *Am J Hum Genet* **94**, 485-495 (2014). https://doi.org:10.1016/j.ajhg.2014.02.011

30 Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212-216 (2015). https://doi.org:10.1038/nature14465

31 Thiagalingam, S. Epigenetic memory in development and disease: Unraveling the mechanism. *Biochim Biophys Acta Rev Cancer* **1873**, 188349 (2020). https://doi.org:10.1016/j.bbcan.2020.188349

32 Hachiya, T. *et al.* Genome-wide identification of inter-individually variable DNA methylation sites improves the efficacy of epigenetic association studies. *NPJ Genom Med* **2**, 11 (2017). https://doi.org:10.1038/s41525-017-0016-5

33 Yoshida-Tanaka, K. *et al.* Narcolepsy type I-associated DNA methylation and gene expression changes in the human leukocyte antigen region. *Sci Rep* **13**, 10464 (2023). https://doi.org:10.1038/s41598-023-37511-4

34 Li, T. *et al.* Epigenomics and transcriptomics of systemic sclerosis CD4+ T cells reveal long-range dysregulation of key inflammatory pathways mediated by disease-associated susceptibility loci. *Genome Med* **12**, 81 (2020). https://doi.org:10.1186/s13073-020-00779-6

35 Clark, A. D. *et al.* Lymphocyte DNA methylation mediates genetic risk at shared immune-mediated disease loci. *J Allergy Clin Immunol* **145**, 1438-1451 (2020). https://doi.org:10.1016/j.jaci.2019.12.910

36 Martino, D. *et al.* Epigenetic dysregulation of naive CD4+ T-cell activation genes in childhood food allergy. *Nat Commun* **9**, 3308 (2018). https://doi.org:10.1038/s41467-018-05608-4

37 Tserel, L. *et al.* Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. *Sci Rep* **5**, 13107 (2015). https://doi.org:10.1038/srep13107

38 Ventham, N. T. *et al.* Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun* **7**, 13507 (2016). https://doi.org:10.1038/ncomms13507

39 Wang, T. *et al.* Enhancing discoveries of molecular QTL studies with small sample size using summary statistic imputation. *Brief Bioinform* **23** (2022). https://doi.org:10.1093/bib/bbab370

40 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013). https://doi.org:10.1186/gb-2013-14-10-r115

41 Gadd, D. A. *et al.* Epigenetic scores for the circulating proteome as tools for disease prediction. *Elife* **11** (2022). https://doi.org:10.7554/eLife.71802

42 Hillary, R. F. & Marioni, R. E. MethylDetectR: a software for methylation-based health profiling. *Wellcome Open Res* **5**, 283 (2020). https://doi.org:10.12688/wellcomeopenres.16458.2

43 Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50**, 591-602 (2018). https://doi.org:10.1038/s41588-018-0073-4

44 Korbecki, J. *et al.* Hypoxia Alters the Expression of CC Chemokines and CC Chemokine Receptors in a Tumor-A Literature Review. *Int J Mol Sci* **21** (2020). https://doi.org:10.3390/ijms21165647

45 Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715.e1716 (2018). https://doi.org:10.1016/j.cell.2018.10.022

46 Yoshie, O. CCR4 as a Therapeutic Target for Cancer Immunotherapy. *Cancers (Basel)* **13** (2021). https://doi.org:10.3390/cancers13215542

47 Maeda, Y. *et al.* Depletion of central memory CD8+ T cells might impede the antitumor therapeutic effect of Mogamulizumab. *Nat Commun* **12**, 7280 (2021). https://doi.org:10.1038/s41467-021-27574-0

48 Li, Z. H. & Bresnick, A. R. The S100A4 metastasis factor regulates cellular motility via a direct interaction with myosin-IIA. *Cancer Res* **66**, 5173-5180 (2006). https://doi.org:10.1158/0008-5472.CAN-05-3087

49 Shimada, M. *et al.* An approach based on a genome-wide association study reveals candidate loci for narcolepsy. *Hum Genet* **128**, 433-441 (2010). https://doi.org:10.1007/s00439-010-0862-z

50 van Daalen, K. R., Reijneveld, J. F. & Bovenschen, N. Modulation of Inflammation by Extracellular Granzyme A. *Front Immunol* **11**, 931 (2020). https://doi.org:10.3389/fimmu.2020.00931

51 Ng, S. S. *et al.* The NK cell granule protein NKG7 regulates cytotoxic granule exocytosis and inflammation. *Nat Immunol* **21**, 1205-1218 (2020). https://doi.org:10.1038/s41590-020-0758-6

52 Moro-García, M. A., Mayo, J. C., Sainz, R. M. & Alonso-Arias, R. Influence of Inflammation in the Process of T Lymphocyte Differentiation: Proliferative, Metabolic, and Oxidative Changes. *Front Immunol* **9**, 339 (2018). https://doi.org:10.3389/fimmu.2018.00339

53 Quante, T. & Bird, A. Do short, frequent DNA sequence motifs mould the epigenome? *Nat Rev Mol Cell Biol* **17**, 257-262 (2016). https://doi.org:10.1038/nrm.2015.31

908 54 Hurtado, C. *et al.* DNA methylation changes on immune cells in Systemic Lupus
909 Erythematosus. *Autoimmunity* **53**, 114-121 (2020).
910 https://doi.org:10.1080/08916934.2020.1722108
911 55 Ma, Q. *et al.* Specific hypomethylation programs underpin B cell activation in early
912 multiple sclerosis. *Proc Natl Acad Sci U S A* **118** (2021).
913 https://doi.org:10.1073/pnas.2111920118
914 56 American Academy of Sleep Medicine. International Classification of Sleep Disorders-
915 Third Edition. Darien, IL: American Academy of Sleep Medicine (2014).
916 57 Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**,
917 428-430 (2014). https://doi.org:10.1093/bioinformatics/btt684
918 58 Tian, Y. *et al.* ChAMP: updated methylation analysis pipeline for Illumina BeadChips.
919 *Bioinformatics* **33**, 3982-3984 (2017). https://doi.org:10.1093/bioinformatics/btx513
920 59 Butcher, L. M. & Beck, S. Probe Lasso: a novel method to rope in differentially
921 methylated regions with 450K DNA methylation data. *Methods* **72**, 21-28 (2015).
922 https://doi.org:10.1016/j.ymeth.2014.10.036
923 60 Heiss, J. A. & Just, A. C. Improved filtering of DNA methylation microarray data by
924 detection p values and its impact on downstream analyses. *Clin Epigenetics* **11**, 15 (2019).
925 https://doi.org:10.1186/s13148-019-0615-3
926 61 Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and
927 innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22
928 (2017). https://doi.org:10.1093/nar/gkw967
929 62 Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting
930 probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**,
931 189-196 (2013). https://doi.org:10.1093/bioinformatics/bts680
932 63 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression
933 data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
934 https://doi.org:10.1093/biostatistics/kxj037
935 64 Wong, W. F., Kohu, K., Chiba, T., Sato, T. & Satake, M. Interplay of transcription factors
936 in T-cell differentiation and function: the role of Runx. *Immunology* **132**, 157-164 (2011).
937 https://doi.org:10.1111/j.1365-2567.2010.03381.x
938 65 Kanhere, A. *et al.* T-bet and GATA3 orchestrate Th1 and Th2 differentiation through
939 lineage-specific targeting of distal regulatory elements. *Nat Commun* **3**, 1268 (2012).
940 https://doi.org:10.1038/ncomms2260
941 66 Lutsik, P. *et al.* MeDeCom: discovery and quantification of latent components of
942 heterogeneous methylomes. *Genome Biol* **18**, 55 (2017). https://doi.org:10.1186/s13059-
943 017-1182-6

944 67 Jaffe, A. E. *et al.* Bump hunting to identify differentially methylated regions in
945 epigenetic epidemiology studies. *Int J Epidemiol* **41**, 200-209 (2012).
946 https://doi.org:10.1093/ije/dyr238
947 68 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
948 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-
949 140 (2010). https://doi.org:10.1093/bioinformatics/btp616
950 69 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor
951 RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-
952 4297 (2012). https://doi.org:10.1093/nar/gks042
953 70 Chen, Y., Lun, A. T. & Smyth, G. K. From reads to genes to pathways: differential
954 expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-
955 likelihood pipeline. *F1000Res* **5**, 1438 (2016).
956 https://doi.org:10.12688/f1000research.8987.2
957 71 Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-
958 Generation Reference Panels. *Am J Hum Genet* **103**, 338-348 (2018).
959 https://doi.org:10.1016/j.ajhg.2018.07.015
960 72 Fuse, N. *et al.* Establishment of Integrated Biobank for Precision Medicine and
961 Personalized Healthcare: The Tohoku Medical Megabank Project. *JMA J* **2**, 113-122
962 (2019). https://doi.org:10.31662/jmaj.2019-0014
963 73 Shimada, M. *et al.* Epigenome-wide association study of narcolepsy-affected lateral
964 hypothalamic brains, and overlapping DNA methylation profiles between narcolepsy and
965 multiple sclerosis. *Sleep* **43** (2020). https://doi.org:10.1093/sleep/zsz198
966 74 Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res* **49**, D916-D923 (2021).
967 https://doi.org:10.1093/nar/gkaa1087
968 75 Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large,
969 centralized data repositories. *Nucleic Acids Res* **43**, W589-598 (2015).
970 https://doi.org:10.1093/nar/gkv350
971 76 Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.
972 *Bioinformatics* **28**, 1353-1358 (2012). https://doi.org:10.1093/bioinformatics/bts163
973 77 Hawe, J. S. *et al.* Genetic variation influencing DNA methylation provides insights into
974 molecular mechanisms regulating genomic function. *Nat Genet* **54**, 18-29 (2022).
975 https://doi.org:10.1038/s41588-021-00969-x
976 78 Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize
977 the Relationship between Methylomic Variation, Gene Expression, and Complex Traits.
978 *Am J Hum Genet* **103**, 654-665 (2018). https://doi.org:10.1016/j.ajhg.2018.09.007

79 Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **41**, D377-386 (2013). https://doi.org:10.1093/nar/gks1118

80 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000). https://doi.org:10.1038/75556

81 Consortium, G. O. Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049-1056 (2015). https://doi.org:10.1093/nar/gku1179

82 Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res* **50**, D988-D995 (2022). https://doi.org:10.1093/nar/gkab1049

83 Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020). https://doi.org:10.1038/s41586-020-2493-4

84 Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res* **50**, D20-D26 (2022). https://doi.org:10.1093/nar/gkab1112

Tables

Table 1. DMRs in NT1 in the CD4$^+$ and CD8$^+$ T cell analyses

CHR	Position		Length (bp)	Hyper- and hypo-methylated in NT1	Bumphunter_P		Neighboring genes	Genes whose expression is associated with methylation*	Detected in both CD4$^+$ and CD8$^+$ T cells
	Start	End			Discovery	Replication			
CD4$^+$ T cells									
8	81787438	81788187	749	Hypo	4.7.E-04	5.7.E-03	ZNF704	PAG1_N	
7	27137427	27138974	1547	Hyper	3.5.E-03	2.6.E-04	HOTAIRM1	NA	
10	81044831	81045166	335	Hypo	6.5.E-03	4.5.E-04	ZMIZ1	NA	
3	33700962	33701223	261	Hyper	8.3.E-03	1.0.E-03	CLASP2	CCR4_P;CRTAP_P	●
1	32264223	32264703	480	Hypo	9.2.E-03	9.3.E-03	SPOCD1	NA	
17	6899085	6899380	295	Hyper	7.1.E-03	5.1.E-04	ALOX12	NA	
1	159046937	159047177	240	Hypo	1.2.E-02	3.0.E-04	AIM2	PYHIN1_N;AIM2_N;TAGLN2_N;SPTA1_P;KIRREL1_N;DUSP23_N	
8	38831681	38832728	1047	Hyper	1.4.E-02	1.1.E-03	HTRA4	NA	
18	77723195	77723602	407	Hypo	1.4.E-02	4.6.E-04	HSBP1L1	NA	
3	1.84E+08	1.84E+08	677	Hypo	1.5.E-02	6.5.E-03	EIF4G1	NA	
17	27048437	27049557	1120	Hypo	1.5.E-02	1.9.E-03	RPL23A;SNORD42A;SNORD4B;SNORD4A	SDF2_N;ALDOC_N;SPAG5_N;CORO6_N;TLCD1_P;SEZ6_P;TMEM199_P;PIPOX_P;SEZ6_P;ABHD15_N;NLK_P	
20	47896888	47897451	563	Hypo	1.7.E-02	1.1.E-04	MIR1259;SNORD12B;C20orf199;SNORD12	B4GALT5_N;PREX1_N;ZNFX1_P	●
2	74727000	74727321	321	Hyper	1.6.E-02	1.2.E-04	LBX2	MRPL53_N	

17	34207332	34207663	331	Hypo	2.1.E-02	1.3.E-03	CCL5	SLFN11_N;ENSG00000267648_P;ENSG00000141150_N;ENSG00000267102_P;PIGW_P	●
11	1892038	1892307	269	Hypo	1.7.E-02	8.2.E-04	LSP1	NA	
CD8$^+$ T cells									
20	47896448	47897451	1003	Hypo	6.7.E-03	2.7.E-03	C20orf199;SNORD12B;SNORD12;MIR1259	B4GALT5_N;PREX1_N;PEDS1_N;STAU1_N;UBE2V1_N;ZNFX1_P	●
6	39758343	39758998	655	Hypo	8.6.E-03	1.3.E-03	DAAM2	NA	
22	31318179	31318546	367	Hypo	9.8.E-03	2.1.E-03	C22orf27	NA	
17	34207332	34208012	680	Hypo	1.1.E-02	1.9.E-03	CCL5	SLFN11_N;ENSG00000267648_P;ENSG00000267102_P;ENSG00000141150_N;PIGW_P	●
3	33701101	33701223	122	Hyper	1.3.E-02	1.5.E-02	CLASP2	CCR4_P;CRTAP_P	●

* Genes for which methylation sites correlated with expression were identified based on Methylation-Expression information from the integrated analysis of methylation and expression data utilizing data from previous studies. '_N' and '_P' indicate a negative and positive correlation, respectively, between the methylation rate and gene expression.

Table 2. The proportion of hypomethylated sites among the NT1-related methylation sites in CD4$^+$ T cells

Discovery set	Replicated with $P < 0.05$*		
Association test	Number of methylation sites		*P*-value**
P-value	All	Hypomethylated in NT1 (%)	
1.00E-05	5	5 (100%)	6.42E-01
0.0001	45	44 (97.8%)	3.24E-02
0.001	364	359 (98.6%)	8.55E-11
0.01	3332	3,259 (97.8%)	5.51E-83
0.05	14,151	13,728 (97.0%)	<1E−308
Analyzed CpG sites***			
	735,841	448,555 (61.0%)	

* Replicated probes only include probes for which the direction of methylation was consistent in both the discovery set and the replication set. In other words, we targeted probes that are hypomethylated at NT1 in the discovery set and are also hypomethylated at NT1 in the replication set, as well as those that are hypermethylated at NT1 in the discovery set and are also hypermethylated at NT1 in the replication set.

** Hypothesis testing for the difference in the population proportions.

*** Probes located in the *HLA* region were excluded.

Table 3. The proportion of hypomethylated sites among the NT1-related methylation sites in CD8$^+$ T cells

Discovery set	Replicated with $P < 0.05$*		
Association test	Number of methylation sites		*P*-value**
P-value	All	Hypomethylated in NT1 (%)	
1.00E-05	4	4 (100%)	7.47E-01
0.0001	23	21 (91.3%)	2.53E-01
0.001	147	123 (83.7%)	1.55E-02
0.01	1,116	897 (80.4%)	5.44E-09
0.05	4,650	3,749 (80.6%)	1.73E-33
Analyzed CpG sites***			
	735,841	455,080 (61.8%)	

* Replicated probes only include probes for which the direction of methylation was consistent in both the discovery set and the replication set.

** Hypothesis testing for the difference in the population proportions.

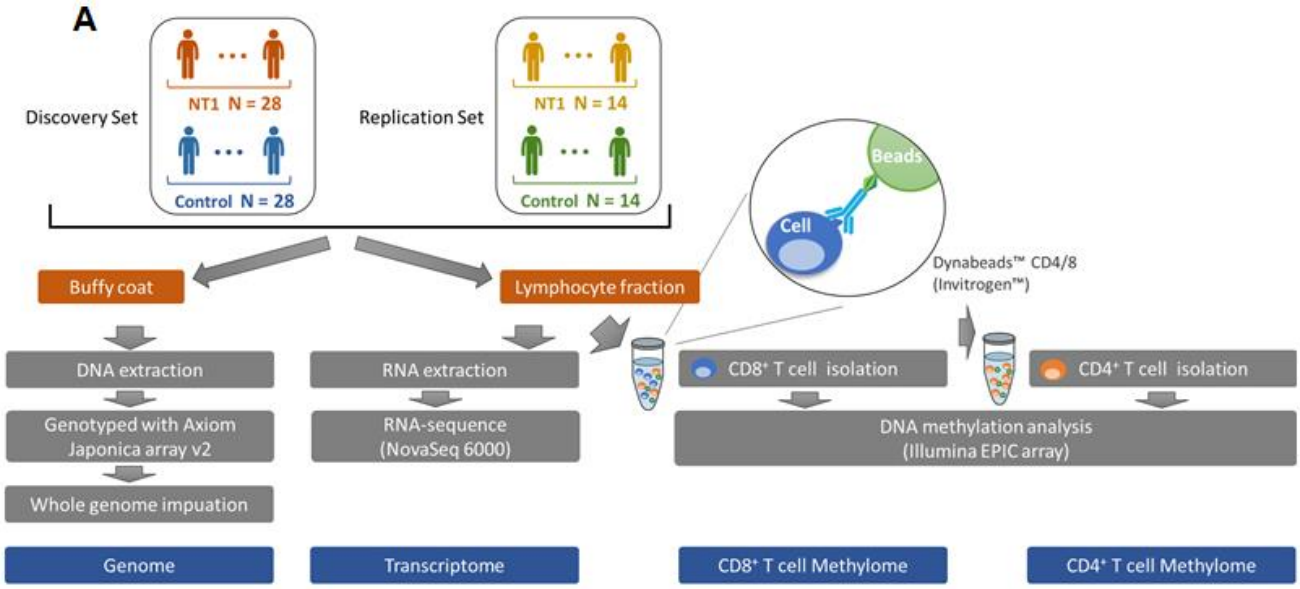
*** Probes located in the *HLA* region were excluded.

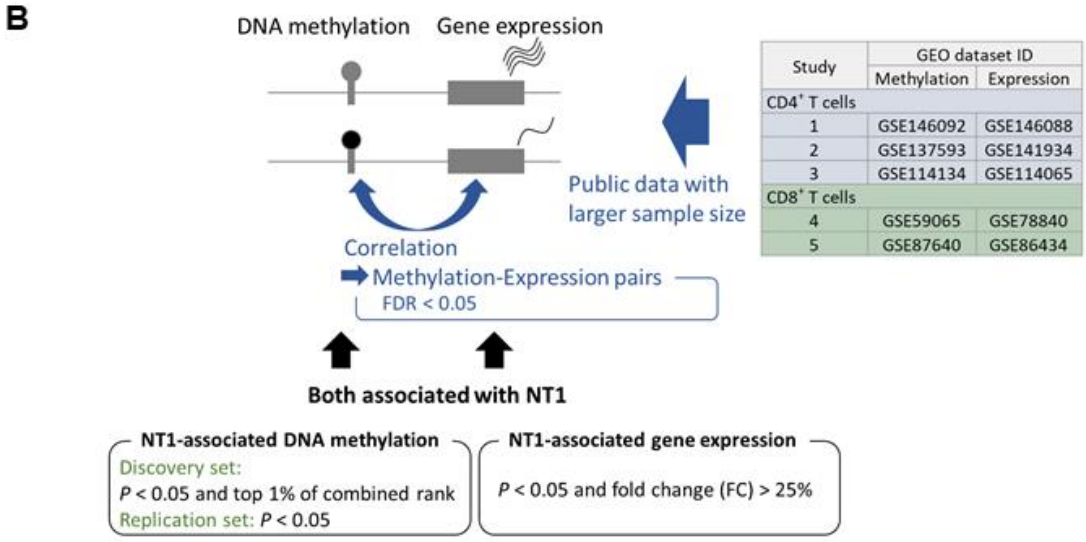
Table 4. NT1-associated hypomethylation sites are abundant in solo-WCGW

	Rate in all analyzed CpGs	CD4$^+$ T cells					CD8$^+$ T cells				
		Rate in hypomethylated NT1 CpGs	*P*-value	OR	95%CI		Rate in hypomethylated NT1 CpGs	*P*-value	OR	95%CI	
					Lower	Upper				Lower	Upper
Solo-WCGW	0.11	0.27	9.87E-194	3.069	2.839	3.317	0.21	1.46E-22	2.199	1.872	2.584
Solo	0.34	0.55	1.12E-150	2.446	2.283	2.622	0.52	9.64E-32	2.154	1.89	2.456
WCGW	0.23	0.45	2.85E-198	2.77	2.585	2.969	0.38	1.89E-29	2.136	1.867	2.444

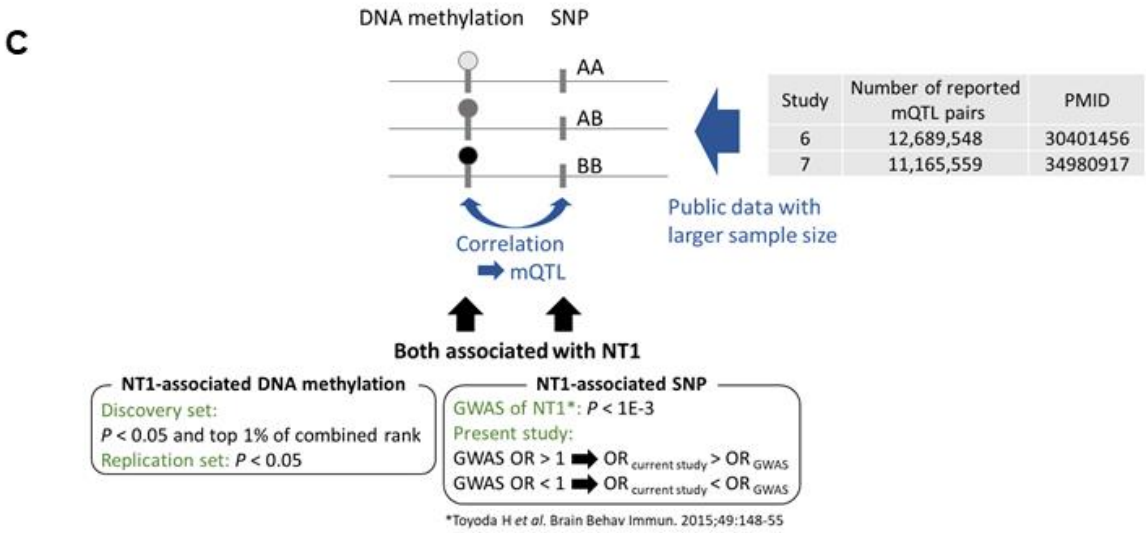
Abbreviation: OR, odds ratio

1017 **Figures**





Study	GEO dataset ID	
	Methylation	Expression
CD4⁺ T cells		
1	GSE146092	GSE146088
2	GSE137593	GSE141934
3	GSE114134	GSE114065
CD8⁺ T cells		
4	GSE59065	GSE78840
5	GSE87640	GSE86434



Study	Number of reported mQTL pairs	PMID
6	12,689,548	30401456
7	11,165,559	34980917

1018 **Figure 1. Overview of the study. A.** Study design and graphical abstract for multiomics
1019 analysis. The discovery set consists of 28 cases of NT1 and 28 controls, while the replication
1020 set consists of 14 cases of NT1 and 14 controls. From each sample, the buffy coat and leukocyte
1021 fractions were collected. DNA was extracted from the buffy coat and genotyped using an array.
1022 From one of the two divided leukocyte fractions, RNA was first extracted, then RNA
1023 sequencing was performed to obtain transcriptome data. Using the remaining leukocyte
1024 fraction, $CD8^+$ T cells were isolated using beads, followed by the isolation of $CD4^+$ T cells.
1025 DNA was extracted from each of these isolated T cell types, and the DNA methylation rate was
1026 measured using an array. **B.** Methodology of the integrative analysis of methylome and
1027 transcriptome data. Initially, we identified pairs of methylation rate and gene expression that
1028 correlated with each other (Methylation-Expression pairs) using data from previous studies
1029 with larger sample sizes. Among the identified pairs, we then explored pairs where both the
1030 methylation rate and gene expression were associated with NT1 in our EWAS and RNA-seq.
1031 **C.** Methodology of the integrative analysis of methylome and genome data. From among the
1032 mQTLs detected in previous studies with larger sample sizes, we searched for mQTLs that had
1033 both a methylation rate and genotype that were associated with NT1 in our EWAS and GWAS.

1034

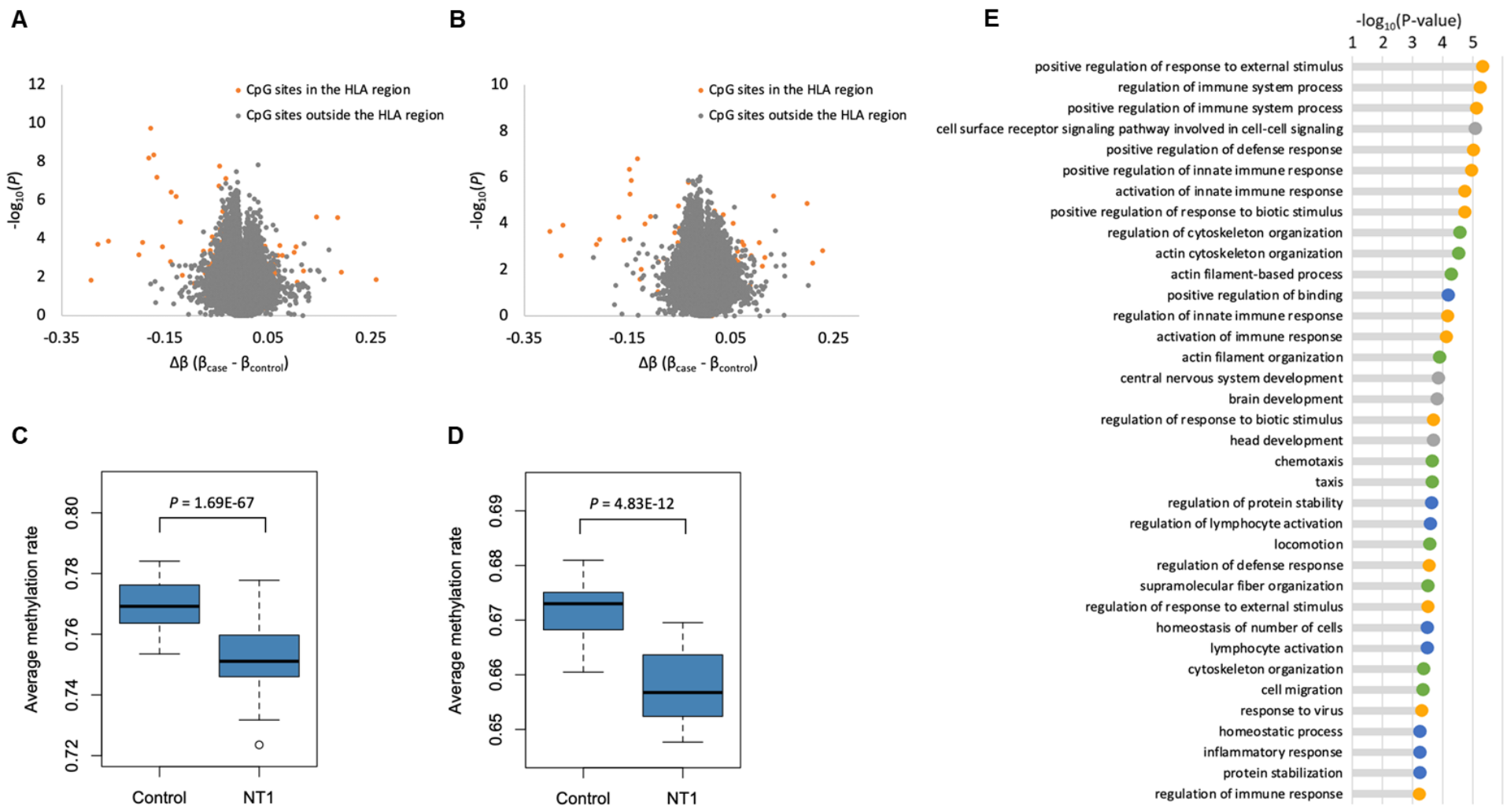

A
CpG sites in the HLA region
CpG sites outside the HLA region
-log10(P)
Δβ (βcase - βcontrol)
B
CpG sites in the HLA region
CpG sites outside the HLA region
-log10(P)
Δβ (βcase - βcontrol)
C
P = 1.69E-67
Average methylation rate
Control
NT1
D
P = 4.83E-12
Average methylation rate
Control
NT1
E
-log10(P-value)
positive regulation of response to external stimulus
regulation of immune system process
positive regulation of immune system process
cell surface receptor signaling pathway involved in cell-cell signaling
positive regulation of defense response
positive regulation of innate immune response
activation of innate immune response
positive regulation of response to biotic stimulus
regulation of cytoskeleton organization
actin cytoskeleton organization
actin filament-based process
positive regulation of binding
regulation of innate immune response
activation of immune response
actin filament organization
central nervous system development
brain development
regulation of response to biotic stimulus
head development
chemotaxis
taxis
regulation of protein stability
regulation of lymphocyte activation
locomotion
regulation of defense response
supramolecular fiber organization
regulation of response to external stimulus
homeostasis of number of cells
lymphocyte activation
cytoskeleton organization
cell migration
response to virus
homeostatic process
inflammatory response
protein stabilization
regulation of immune response

Figure 2. Results of the association analysis in $CD4^+$ and $CD8^+$ T cells. A. Volcano plot of the DMP analysis in the $CD4^+$ T cells. **B.** Volcano plot of the DMP analysis in the $CD8^+$ T cells. **C.** Q-Q plot of the DMP analysis in the $CD4^+$ T cells. We plotted the methylation sites separately, categorizing them into those that are hypomethylated and those that are hypermethylated in NT1. **D.** Q-Q plot of the DMP analysis in the $CD8^+$ T cells. Similar to the $CD4^+$ T cell analysis, we plotted the results of the methylation sites separately as hypomethylated or hypermethylated in NT1. **E.** Results of the pathway analysis using genes suggested to be associated with NT1 in $CD4^+$ T cells. For easier interpretation, pathways were clustered into four groups using the k-means method. The groups were color-coded as follows: yellow, pathways related to the immune response; green, pathways related to cell migration; gray, neural pathways; and blue, others.

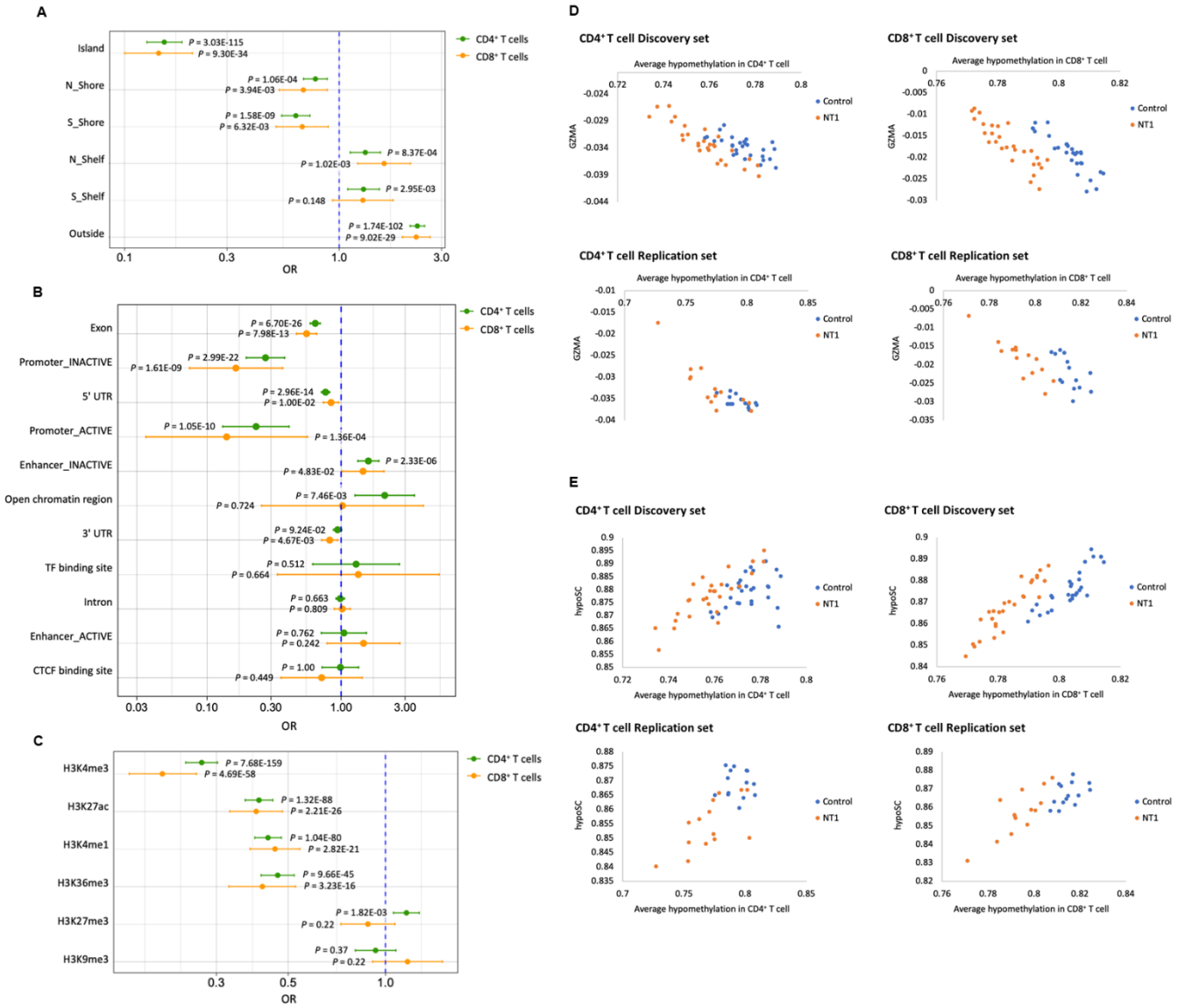


Figure 3. Regions with a high frequency of NT1-related hypomethylation sites and indices that showed a correlation with the degree of hypomethylation. A. The relationship between NT1-related hypomethylation sites and CpG islands. **B.** The relationship between NT1-related hypomethylation sites and gene regions. Regarding promoters and enhancers, categories other than "active" and "inactive" have been omitted from the figure, and the detailed results are described in Table S8. **C.** The relationship between NT1-related hypomethylation sites and regions marked by histone modifications. **D**. Correlation between the estimated values of GZMA and the degree of hypomethylation (adjusted for age). **E.** Correlation between hypoSC and the degree of hypomethylation (adjusted for age).