5-Hydroxymethylcytosine Remodeling Precedes Lineage Specification during Differentiation of Human CD4+ T Cells

Graphical Abstract

Highlights

- 5hmC remodeling is widespread during human CD4+ T cell differentiation
- Early 5hmC gains predict loss of DNA methylation in differentiated cells
- Early 5hmC remodeling in vitro predicts loss of DNA methylation in vivo
- 5hmC loci are enriched for functional T cell disease-associated genetic variants

Authors
Colm E. Nestor, Antonio Lentini, Cathrine Hägg Nilsson, ..., Helmut Laumen, Huan Zhang, Mikael Benson

Correspondence
colm.nestor@liu.se (C.E.N.), mikael.benson@liu.se (M.B.)

In Brief
Nestor et al. reveal widespread 5hmC-mediated DNA de-methylation during in vitro differentiation of human CD4+ T cells. They find that regions undergoing 5hmC remodeling are enriched for disease-associated regulatory regions.

Accession Numbers
E-MTAB-4685
E-MTAB-4686
E-MTAB-4687
E-MTAB-4688
E-MTAB-4689

Nestor et al., 2016, Cell Reports 16, 559–570
July 12, 2016 © 2016 The Author(s).
http://dx.doi.org/10.1016/j.celrep.2016.05.091
SUMMARY

5-methylcytosine (5mC) is converted to 5-hydroxymethylcytosine (5hmC) by the TET family of enzymes as part of a recently discovered active DNA demethylation pathway. 5hmC plays important roles in regulation of gene expression and differentiation and has been implicated in T cell malignancies and autoimmunity. Here, we report early and widespread 5mC/5hmC remodeling during human CD4+ T cell differentiation ex vivo at genes and cell-specific enhancers with known T cell function. We observe similar DNA demethylation in CD4+ T cell differentiation in vivo, indicating that early remodeling events persist long term in differentiated cells. Underscoring their important function, 5hmC loci were highly enriched for genetic variants associated with T cell diseases and T-cell-specific chromosomal interactions. Extensive functional validation of 22 risk variants revealed potentially pathogenic mechanisms in diabetes and multiple sclerosis. Our results support 5hmC-mediated DNA demethylation as a key component of CD4+ T cell biology in humans, with important implications for gene regulation and lineage commitment.

INTRODUCTION

Differentiation of CD4+ T cells into effector or regulatory subtypes is critical to adaptive immunity. Upon contact with antigen, T cells differentiate into various T helper (Th) cell subsets, such as Th1, Th2, Th17, or regulatory T (Treg) cells (Yamane and Paul, 2013), which mediate or inhibit immune responses. Inappropriate CD4+ T cell differentiation is associated with several autoimmune and inflammatory diseases, including rheumatoid arthritis (RA), psoriasis, allergy, asthma, multiple sclerosis (MS), and type 1 diabetes (Gustafsson et al., 2015; Licona-Limon et al., 2013; Wahren-Herlenius and Doern, 2013). The lack of a strong genetic component and increasing prevalence of these diseases suggests an epigenetic contribution to their pathogenesis, and changes in T cell DNA methylation patterns have been reported in MS, allergy, and RA (Graves et al., 2013; Liu et al., 2013; Nestor et al., 2014a).

Appropriate differentiation of Th cell subsets requires widespread remodeling of the T cell epigenome, including DNA demethylation of key effector genes, such as Ifng and Il4, Il5, and Il13 for Th1 and Th2, respectively (Janson et al., 2011; Lee et al., 2006). 5-hydroxymethylcytosine (5hmC) recently was discovered to be highly abundant in the human genome and generated by hydroxylation of 5-methylcytosine (5mC) by members of the Ten-Eleven-Translocation (TET1/2/3) family of enzymes (Tahiliani et al., 2009). 5hmC subsequently can be resolved to unmodified cytosine, completing the process of DNA demethylation (Figure S1A). Significantly, TET
loss-of-function mutations, with have been identified in several hematological malignancies, with the highest frequency in adult CD4+ T cell leukemias (Kalender Atak et al., 2012; Lemonnier et al., 2012). Moreover, TET2 knockout mice exhibited impaired differentiation of hematopoietic stem cells and developed autoimmune phenotypes (Ichiyama et al., 2015; Ko et al., 2011; Li et al., 2011; Ko et al., 2015). Despite the valuable insights into the role of TET-5hmC during differentiation of mammalian CD4+ T cells obtained from mouse models (Ichiyama et al., 2015; Ko et al., 2011; Tsagaratou et al., 2014; Yang et al., 2015), little is known about the importance of DNA de-methylation in human CD4+ T cell differentiation and its contribution to the pathogenesis of complex immune diseases.

We generated genome-wide 5hmC, 5mC, and gene expression profiles during early and late stages of human CD4+ T cell differentiation ex vivo. Changes in 5hmC were widespread during both activation and differentiation of CD4+ T cells, occurred at a variety of activating and repressive regulatory elements, and coincided with tight regulation of TET gene expression. Significantly, all early 5hmC and 5mC remodeling occurred in the complete absence of replication, suggesting an active enzymatic remodeling mechanism. Using genetic overexpression, we showed that tight regulation of TET levels was required for appropriate expression of key lineage-specific transcription factors and cytokines. We confirmed these findings in vivo by transcriptional and epigenetic profiling of human naive CD4+ (NT) T cells and central memory (TCM) and effector memory (TEM) T cells. Supporting the disease relevance of 5hmC-mediated DNA de-methylation, loci gaining 5hmC during early T cell differentiation were highly enriched for variants associated with T cell-related diseases at a diversity of cis and trans gene-regulatory elements. Moreover, these regions also were enriched for T cell-specific chromosomal interactions, supporting their importance in T cell biology. We undertook further functional characterization of the effects of over 20 predicted regulatory variants on the level of DNA-protein interactions, and we reveal potential pathogenic mechanisms in diabetes and MS. Our results support 5hmC-mediated DNA de-methylation as a key component of CD4+ T cell biology in humans and 5hmC profiling as an effective approach for the identification of regulatory genetic variants in complex immune disease.

RESULTS

5hmC Remodeling during CD4+ T Cell Differentiation Occurs in the Absence of Replication and Is Enriched at Key Regulatory Genes

To dissect the role of DNA de-methylation in human CD4+ T cell function, we leveraged the ability to differentiate pure human NT cells into Th cell subsets in vitro (Figure 1A). Appropriate differentiation into Th1 and Th2 lineages was confirmed by gene expression microarray and qRT-PCR of key lineage-specific genes (Figures S1B and S1C; Table S1). In vitro differentiation (polarization) of NT cells into Th1 and Th2 lineages resulted in a >5-fold loss (p < 0.05, t test) of global 5hmC levels (Figure 1B). A small but consistent 10%–15% loss of global 5mC levels also was observed after only 1 day of in vitro polarization before the onset of DNA replication (Figures 1B, S1D, and S1E), suggesting DNA replication-independent remodeling of 5hmC. To investigate locus-specific remodeling of 5hmC, we combined 5hmC DNA immunoprecipitation with massively parallel sequencing (hydroxymethylated DNA immunoprecipitation [hMeDIP]-seq) to generate genome-wide 5hmC profiles in a time series of human CD4+ T cells polarized toward Th1 and Th2 lineages. 5hmC profiles were generated for primary NT cells and at early (1-day) and late (5-day) time points during polarization, allowing identification of both early replication-independent changes as well as late subset-specific changes in 5hmC. For each condition, 35–60 million paired reads were generated, resulting in one billion paired reads that were uniquely mapped to the human genome (hg19) (Table S2). The 5hmC profiles showed characteristic enrichment in gene bodies and depletion at transcription start sites (TSSs), as previously reported in other cell types (Figure 1C), as well as a clear association with actively transcribed genes (Figure 1D) (Nestor et al., 2012; Song et al., 2011). Similar to results reported in mouse CD4+ T cells, the distribution of 5hmC enrichment was significantly overrepresented in genic regions in all conditions (Figure 1E; Table S2) (Ichiyama et al., 2015; Tsagaratou et al., 2014).

To understand the biological relevance of these changes, we mapped the 1,000 largest gains and losses of 5hmC to their nearest gene and subjected these to gene ontology (GO) enrichment analysis. Strikingly, whereas regions losing 5hmC were not enriched in any biological processes, regions gaining 5hmC at day 1 in either Th1 or Th2 cells were highly significantly enriched for genes associated with T cell activation (adjusted p < 1 × 10−5; Figure 1F; data not shown), including IFNG, TBX21, FOXP3, ZAP70, IL2RA, IL7R, IRF4, CD4, AIM2, CCR2, CCR5, IL1R2, IL26, and IL32R (Figure 1G; data not shown). Globally, large-scale remodeling of 5hmC had occurred by day 1, with over 10,000 and 5,000 regions significantly gaining or losing 5hmC, respectively (Figure S1F; Table S2). The scale of change was far greater in regions gaining 5hmC (Figure S1G).

Taken together, these results reveal genome-wide, replication-independent reprogramming of 5hmC during CD4+ T cell differentiation at genes key to appropriate T cell activation and differentiation.

Lineage Specification of Human CD4+ T Cells Is Preceded by 5hmC-Mediated DNA De-methylation of Gene-Regulatory Elements

To relate 5hmC remodeling during differentiation to changes in DNA methylation (5mC), we subjected the same in-vitro-differentiated human Th1 and Th2 cells to DNA methylation profiling using Infinium 450K methylation arrays. Although not genome-wide, these arrays provide quantitative, base resolution methylation measurements at ∼450,000 CpG sites throughout the genome, allowing accurate detection of even small (>3%) changes in absolute methylation levels. We observed a clear bias toward loss of 5mC in both Th1 and Th2 differentiated cells, with more pronounced changes occurring by day 5 (Figure 2A). However, some changes in DNA methylation also were observed at day 1 (123 CpGs > 10% loss, p < 0.05), in the complete absence of DNA replication (Figure S2A; Table S3), suggesting that such changes are active and not secondary to DNA replication. Importantly, regions gaining 5hmC at day 1 of differentiation...
in both Th1 and Th2 cells were highly significantly enriched for CpGs losing 5mC at day 5 ($p < 0.0001$, Fisher’s exact test) (Figures 2B and S2B), suggesting that DNA de-methylation at these loci occurs primarily via a 5hmC-mediated process. Indeed, loci showing the greatest loss of 5mC during Th1 and Th2 differentiation were also those showing the greatest gains in 5hmC after only 1 day of polarization (Figure 2C). Finally, genes showing loss of DNA methylation (>30% loss of 5mC) in their promoters during differentiation were enriched for functional terms related to the immune response (Figure S2C), further indicating that DNA de-methylation occurs in a targeted manner. Taken together, these findings suggest a model whereby loci to be de-methylated in differentiated T cell subtypes first undergo enzymatic hydroxylation of 5mC to 5hmC, followed by both replication-independent excision and replication-dependent dilution of 5hmC during successive rounds of DNA replication.
As the majority (>95%) of loci undergoing 5hmC remodeling during early CD4⁺ T cell differentiation occurred outside annotated gene promoters, we hypothesized that these loci might represent gene-regulatory elements. Indeed, the TET enzymes interact with numerous chromatin modifiers (HDAC1, HDAC2, EZH2, and SIN3A) and co-repressor complexes (NuRD), which may serve to target TET methylcytosine dioxygenase activity to regulatory elements within the genome (Delatte et al., 2014), and elevated 5hmC levels have been observed at active enhancers in several systems (Lu et al., 2014; Tsagaratou et al., 2014). Using published chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq) data, we analyzed the association...
between 5mC/5hmC remodeling in CD4+ T cells and the enhancer-associated histone modifications, H3K4me1 and H3K27ac (Hawkins et al., 2013) (Figure 2D). Indeed, Th1- and Th2-specific active enhancer elements marked by H3K27ac (Hawkins et al., 2013) were highly enriched for regions showing 5mC loss during differentiation (~15-fold, p < 0.001) (Figures 2E–2G).

Ectopic Expression of TET1 Results in Dysregulation of Key Cytokine and Chemokine Genes during Differentiation

Expression profiling by qPCR showed that TET1, in contrast to TET2 and TET3, was highly expressed in T cells, as well as thymus compared to other human tissues and immune cells (Figures 3A and S3A). Moreover, whereas the absolute levels of TET1 were lower than those of TET2 and TET3 during T cell differentiation, TET1 alone underwent rapid and stable downregulation during early T cell activation (>10-fold reduction; p < 0.01, t test) (Figures 3B, 3C, and S3B). Indeed, TET1 silencing was observed as early as 6 hr after the initiation of differentiation (Figure S3B). Unlike TET2 and TET3, TET1 has been shown to bind to Polycomb target gene promoters and associate with chromatin repressors, thereby having a role in direct transcriptional repression unrelated to its enzymatic activity (Williams et al., 2011; Wu et al., 2011). Thus, we sought to dissect the functional role of TET1 during early T cell differentiation by ectopic expression of full-length human TET1 (TET1fl) (Tahiliani et al., 2009), the catalytic domain of TET1 (TET1cd), or a mutated catalytic domain of TET1 (TET1mut) (Guo et al., 2011) (Figures 3D, S3C, and S3D). Notably, overexpression of TET1fl led to significant dysregulation of chemokine and cytokine gene mRNA levels (log2 FC > 0.5), including several key regulators of Th1/Th2 differentiation, IFNG, IL12RB2, HAVCR2, GATA3, and IL5. (Figures 3E and S3E). Far fewer changes in gene expression were observed with TET1cd (Figures 3F and S3F; Table S4), the majority of which were associated with calcium ion channel activity (Figure S3F). These findings suggest that TET1 might act as a direct transcriptional repressor during T cell differentiation. However, as the transcriptional programs of each Th cell subset generally inhibit those of the other lineages, future studies are warranted to fully elucidate the functional role of TET1. Nevertheless, the results clearly support an important role for TET1 in the regulation of key genes during lineage-specification of Th cell subsets.

5hmC Remodeling during Early Differentiation of Human CD4+ T Cells Predicts Loss of 5mC in CD4+ Memory T Cells In Vivo

We and other have shown previously that adaptation of primary mammalian cells to culture can affect the genomic distribution of 5mC (Nestor et al., 2012, 2015). Thus, having determined the DNA methylation dynamics of human T cell differentiation in vitro, we sought to establish if similar changes occurred in vivo. After activation and differentiation of NT cells, a small proportion of cells remain as long-lived memory cell populations, which can be sub-divided into TCM and TEM subsets based on their function and homing capacity (Sallusto et al., 2004). We isolated primary human NT (CD4+CD45RO−CCR7+), TCM (CD4+CD45RO−CD45RA−CCR7+), and TEM (CD4+CD45RO-CCR7−). We used a combination of qPCR and mass spectrometry to measure 5mC and 5hmC levels in NT and TEM cells. We found that the ratio of 5hmC to 5mC was higher in TEM cells compared to NT cells, suggesting that 5hmC levels are increased during T cell differentiation. This finding is consistent with previous studies that have shown that 5hmC levels are increased in memory T cells compared to naive T cells (Williams et al., 2011; Wu et al., 2011). Therefore, 5hmC remodeling during early T cell differentiation predicts loss of 5mC in CD4+ memory T cells in vivo.
Figure 4. Early DNA Methylation Remodeling Persists in CD4+ Memory T Cells In Vivo
(A) Barplot of TET gene expression in primary NT cells, central memory (TCM) cells, and effector memory (TEM) cells. Gene expression of TET1/2/3 was measured by qPCR.
(B) Barplot of global 5hmC content was measured by immuno-dot blot using a 5hmC antibody in primary NT, TCM, and TEM cells.
(C) Heatmap of gene expression in primary NT, TCM, and TEM cells showing subset-specific gene signatures. Gene expression was measured by microarray.
(D) Unsupervised hierarchical clustering of DNA methylation (5mC) in NT, TCM, and TEM cells measured by 450K methylation array. Significance was calculated by bootstrap resampling.

(legend continued on next page)
for Disease-Associated Genetic Variants in vivo. Zhang, 2014), 5hmC in CD4+ T cells marked several different (F) Correlation between 5mC and gene expression in T EM cells calculated using Spearman’s rank-correlation coefficient. 5mC was measured by 450k array and gene expression was measured by microarray.

Next we generated gene expression and genome-wide DNA methylation profiles for NT, T CM, and T EM cells. Gene expression of T CM and T EM cells showed distinct profiles, with T EM cells expressing high levels of effector molecules, such as IL4, IFNG, and CSF2, while T CM cells showed an intermediate profile, both in expression levels and number of genes up- or downregulated (Figures 4C, S4B, and S4C; Table S5), consistent with a linear differentiation program of NT → T CM → T EM. Interestingly, DNA methylation profiles could clearly separate NT, T CM, and T EM cells, indicating that each subset is epigenetically distinct (Figures 4D and 44D). Furthermore, T EM cells showed extensive genome-wide de-methylation, exceeding that of T CM cells (Figure 4E), consistent with the notion that T EM cells are more terminally differentiated. As expected, DNA methylation changes in gene promoters were correlated with gene expression ($\rho = -0.43$, $p = 5.9 \times 10^{-06}$) (Figure 4F).

Having characterized methylome- and transcriptome dynamics in memory T cells, we sought to relate these changes to early hydroxymethylome reprogramming during T cell polarization. Consistent with our findings in polarized T cells, regions gaining 5mC after 1 day of polarization exhibited extensive de-methylation in memory cells ($p = 6.20 \times 10^{-78}$; Fisher’s exact test) (Figures 4G–4I). As 5hmC and 5mC cannot be distinguished by conventional bisulfite conversion (Nestor et al., 2014b), we performed 5hmC/5mC-specific qPCR for selected loci undergoing 5mC changes in NT, T CM, and T EM cells (Figures S4E and S4F; Table S6). Thus, early methylome-reprogramming events initiated by 5mC are maintained long term in memory T cells in vivo.

5hmC Remodeling Marks Regulatory Regions Enriched for Disease-Associated Genetic Variants

Consistent with previous reports in other cell types (Wu and Zhang, 2014), 5mC in CD4+ T cells marked several different types of gene-regulatory elements (such as enhancers, promoters, and gene bodies). Using a Hidden Markov model together with published genome-wide datasets for several epigenetic marks in primary human NT cells (Bernstein et al., 2010; Song and Chen, 2015) (Table S7), we observed that 5hmC occupied a unique position in the genome, associating with both repressive and activating states, supporting the property of 5hmC as a general marker of gene-regulatory activity (Figure 5A). Since the majority of disease-associated variants are found in non-coding regions, we compared the frequency of all disease-associated variants reported at the time of analysis ($N_{\text{lead Variant}} = 73,196$; $p < 1 \times 10^{-5}$, GWASdb2) (Li et al., 2012) and those in high linkage disequilibrium (LD) ($N_{\text{VARIANT}} = 600,320$, $r^2 > 0.8$, 1000 Genomes Project, phase 3) in regions gaining or losing 5hmC at day 1 of in vitro polarization, resulting in a total of 1,560 variants (Table S8). Regions gaining 5hmC (hereafter referred to as 5hmC regions) were significantly enriched ($p < 0.01$) for disease-associated variants in 19 diseases, where six of the top ten diseases were autoimmune, whereas regions losing 5hmC showed no significant enrichment for CD4+ T cell diseases (Figure 5B). Strikingly, regions gaining 5hmC were twice as likely (odds ratio [OR] 2.6, $p < 0.0001$, Fisher’s exact test) to overlap a disease-associated variant than cell-type-specific enhancers (Figure 5C) (Hawkins et al., 2013).

Previously, 5hmC has been associated with transcription factor-binding sites (TFBSs) (Yu et al., 2012), suggesting that variants in 5hmC regions may modulate gene expression by disrupting TF binding. We tested all 1,560 5hmC variants using phylogenetic module complexity analysis (PMCA), leveraging the conservation of co-occurring TFBS patterns within gene-regulatory modules, as previously described by us (Claussnitzer et al., 2014). Interestingly, for the variants in 5hmC regions, we found 49.4% predicted to be regulatory compared to our previous findings of 33.3% in a random set of variants (Claussnitzer et al., 2014) (Table S8). To experimentally evaluate the effects of identified non-coding variants on regulatory protein binding, we performed electrophoretic mobility shift assays (EMSAs) using nuclear protein extracts from both transformed and primary, trait-related human CD4+ T cells, with probes for the risk and non-risk alleles of 22 variants (44 alleles) (Table S9). First, we tested variants associated with Crohn disease and MS, as these diseases showed the highest enrichment for variants in 5hmC regions (Figure 5B; Table S8). We also tested variants associated with non-enriched T cell-associated diseases (such as allergy) and other non-T cell-associated traits (obesity and metabolite levels) with high TFBS module conservation (Figure 5B; Table S8). We found an allele-specific shift in 16 of 22 tested variants when using protein extracts from the Jurkat T cell line (Figures 5D and S5A), supporting that 5hmC can aid in the identification of regulatory variants, potentially contributing to disease pathophysiology. Importantly, variant-induced alterations on regulator binding could be replicated by using...
Figure 5. 5hmC Regions Are Highly Enriched for Disease-Associated Variants

(A) ChromHMM heatmap shows enrichment and colocalization of 5hmC with other epigenetic marks in NT cells.

(B) Enrichment of disease-associated variants in peaks gaining (left) and losing (right) 5hmC after 1 day of CD4+ T cell polarization is shown.

(legend continued on next page)
nuclear protein from primary human CD4+ T cells isolated from healthy individuals (Figure 5E).

As none of the risk variants affecting regulator binding were located within gene promoters or gene-coding regions, we assessed how risk variants might modulate gene expression via physical interaction with distal gene-regulatory elements. Using chromatin interaction data (capture Hi-C [cHi-C]) for a human CD4+ T cell line (Martin et al., 2015), we found that disease-associated variants located in 5hmC regions were significantly enriched (bootstrap p < 1 × 10^-4) for long-range chromatin interactions compared to all disease-associated variants (Figure 5F). Interestingly, enrichment of chromatin interactions for disease-associated variants in 5hmC regions was greater than that observed for T cell-specific enhancers (Hawkins et al., 2013), even though, not surprisingly, enhancer regions alone had a significantly higher number of chromatin interactions compared to 5hmC regions (Figure 5F; data not shown). These findings strongly support the power of 5hmC as a precise and efficient marker to identify regulatory disease-associated variants.

To further validate our approach, we examined interactions at the CLEC16A locus, which is associated with several autoimmune diseases (Wellcome Trust Case Control, 2007) (Figure 5G). Of 30 variants in high LD (r^2 > 0.8), five were located in 5hmC regions, two of which were significantly enriched for TFBS module conservation (Table S8). For these two variants, we observed differential protein binding for rs7203150 in both Jurkat and primary human CD4+ cells (Figures 5D and 5E; p = 5.9 × 10^-7/1.8 × 10^-3, respectively, Student’s t test), whereas no allele-specific binding was observed for rs7198004 (Table S8; Figure S5A). The variant rs7203150 is located in intron 19 of the CLEC16A gene, containing many variants associated with multiple autoimmune diseases, and has been shown to physically interact with and regulate the expression of a neighboring gene, DEXI (Davison et al., 2012). The cHi-C data confirmed the highly significant interaction of the rs7203150 region with the DEXI promoter (false discovery rate [FDR] < 5%), as expected, but also revealed other interactions with a downstream gene, RMI2 (Figure 5G). In fact, cHi-C data suggest interactions for numerous genes at the CLEC16A locus, including SOCS1, a negative regulator of cytokine signaling (Diehl et al., 2000); DEXI, a gene of unknown function but proposed as autoimmune candidate gene (Leikfoss et al., 2013); RMI2, a topoisomerase critical for T cell differentiation and a primary immunodeficiency syndrome candidate gene (Mönich et al., 2010); and CLEC16A, recently shown to affect T cell selection, mediate T1D, and modify CD4 single-positive thymocyte reactivity (Schuster et al., 2015) (Figure 5F). Finally, in our T cell data we find supportive co-expression of CLEC16A with SOCS1, DEXI, and RMI2 (Figure 5G, right panel).

Thus, 5hmC remodeling marks cell-type-specific gene-regulatory regions, which may contribute to pathogenic mechanisms in CD4+ T cell-associated diseases.

**DISCUSSION**

Previous studies of 5hmC during T cell differentiation have reported global loss of 5hmC in differentiated Th cell subsets in mouse (Ichiyama et al., 2015; Tsagaratou et al., 2014; Yang et al., 2015). We confirm these findings, but using a time series-profiling approach, we also report widespread 5hmC remodeling during early differentiation, including >10,000 locus-specific gains of 5hmC enriched at genes associated with T cell function. We reveal that 5hmC-mediated DNA de-methylation is a key feature of human CD4+ T cell differentiation and that tight regulation of TET gene expression is critical for appropriate differentiation of Th cell subsets. These DNA de-methylation events also were observed in memory T cells in vivo, and they were enriched for variants associated with CD4+ T cell diseases, including MS, psoriasis, and Crohn disease. Several disease-associated variants affected DNA-protein interactions in primary human CD4+ T cells, revealing potentially pathogenic mechanisms in several autoimmune diseases. As 5hmC profiling can be performed on small amounts (>25 ng) of archived, fragmented genomic DNA, it is ideal for identifying and stratifying potential regulatory variants in rare primary cell types (Taiwo et al., 2012). This is in contrast to chromatin-based enhancer-profiling strategies, which typically require profiling of multiple epigenetic marks in large amounts of fresh material (Hawkins et al., 2013).

Studies in mouse have provided valuable insights into the role of DNA methylation during differentiation of mammalian CD4+ T cells, but less is known about the importance of DNA demethylation in human CD4+ T cell differentiation and its contribution to the pathogenesis of complex immune diseases (Ichiyama et al., 2015; Ko et al., 2011; Tsagaratou et al., 2014; Yang et al., 2015). Our profiling of the methylome (5mC), hydroxymethylome (5hmC), and transcriptome of matched human CD4+ T cells during in vitro differentiation of NT cells into Th1 and Th2 cells revealed widespread remodeling of 5hmC during early T cell activation, and it showed that early gains in 5hmC predicted subsequent loss of 5mC in differentiated Th1 and Th2 cells. Significantly, 5hmC remodeling was enriched in the regulatory regions of genes with known T cell function, including master regulators of lineage specification, such as IFNG, TBX21, and FOXP3. Consistent with an important role for 5hmC-mediated DNA demethylation in T cell differentiation, we found that (1) the TET genes are most highly expressed in T cell-related tissues (CD4+ T cells, CD8+ T cells, and thymus) in humans, (2) CD4+ T cell activation results in dramatic (10-fold) and rapid (6-hr)

(C) Venn diagram shows disease-associated variant localization in 5hmC 1-day gain peaks and CD4+ T cell enhancers.

(D and E) EMSA analysis of protein-DNA binding showing changes in binding upon introduction of disease-associated variants in 5hmC regions using nuclear protein extracts from Jurkat T cell line (D) and primary CD4+ T cells (E). Dashed boxes indicate shifts in binding and arrows indicate shifts not observed in cell line. (F) Barplot of capture Hi-C (cHi-C) interactions overlapping variants in identified 5hmC regions or T cell enhancers. Region sizes were normalized to avoid size bias and p values were calculated using bootstrap resampling (n = 10,000).

(G) Genomic plot of variant rs7203150 located in identified 5hmC region showing interactions with nearby gene promoters (left). Heatmap of co-expression during T cell polarization shows high degree of co-expression between genes interacting with variant rs7203150 (right). cHi-C, capture Hi-C. See also Figure S5.
downregulation of TET1 expression, and (3) dysregulation of TET1 expression disrupts CD4+ T cell differentiation. These observations are consistent with a growing number of studies in Tet knockout mice, which have revealed a pivotal role of TET enzymatic activity in T cell biology (Ichiyama et al., 2015; Ko et al., 2010; Tsagaratou et al., 2014; Yang et al., 2015).

In addition to direct differentiation into effector subtypes, a proportion of activated NT cells become long lived and retain memory of the initial activating signal (Sallusto et al., 1999). Methylome profiling of NT, TCM, and TEM cells ex vivo revealed that the DNA methylation changes observed early during differentiation persist in memory T cells. First, this important observation supports the validity of in vitro CD4+ T cell differentiation as a powerful model system in which to study the epigenetics of human CD4+ T cells. Second, these data suggest that any dysregulation of the DNA de-methylation pathway not only affects immediate response to antigen (differentiation) but also appropriate formation of T cell memory.

The availability of global profiles for several activating and repressive epigenetic marks allowed us to study the epigenetic neighborhood occupied by 5hmC in human CD4+ T cells. We used a recently described Hidden Markov model-based method to identify the distinct chromatin states present in human NT cells (Song and Chen, 2015). Interestingly, instead of simply following the patterning of 5mC, from which it is derived, 5hmC was present in chromatin states composed of activating, repressive, and poised histone modifications (Figure 5A). 5hmC’s presence across so many chromatin states may reflect its property as an intermediate during conversion of transcriptionally repressive 5mC to transcriptionally permissive unmethylated cytosine (Ikon et al., 2014). Alternatively, 5hmC’s association with such a diverse range of histone marks may reflect the growing realization of the TET enzymes as multi-faceted proteins, connecting different layers of epigenetic and transcription control to maintain cell state (Laird et al., 2013). Indeed, the TET enzymes have been reported to directly interact with transcriptional activators, including PU.1, EBF1, and p300, while also interacting with proteins associated with transcriptional repression, including EZH2, SIN3A, HDACs, NuRD, and MeCP2 (Cartron et al., 2013; Delatte et al., 2014).

Regardless of the underlying mechanism, 5hmC’s position as a general but sensitive indicator of chromatin state makes it a powerful tool in the identification of cell-type-specific regulatory elements. Indeed, regions gaining 5hmC during early activation (day 1) were highly enriched for variants associated with several autoimmune diseases, including Crohn disease, MS, celiac disease, myasthenia gravis, and psoriasis. The effect of these potentially regulatory variants on DNA-protein binding was directly tested in primary human CD4+ T cells, and it provides multiple molecular pathways for further investigation. By combining these results with chromatin interaction (CtC) data from a human T cell line, we found that 5hmC variants are enriched for connections between distal regulatory elements, and we identified several interactions, which, if disrupted by risk variants, may have pathogenic consequences in autoimmune disease.

Combined with the ability to perform 5hmC profiling on small amounts of archived DNA, this surprising and provocative finding suggests that 5hmC alone may be a powerful and cost-effective approach for the prioritization of regulatory variants in humans.

In conclusion, 5hmC-mediated DNA de-methylation plays key roles in the differentiation of human CD4+ T cells, marking regions relevant for the pathogenesis of several autoimmune diseases, and it is an effective and accessible approach for identifying causal disease variants in autoimmune disease.

**EXPERIMENTAL PROCEDURES**

**Ethics Statement**

This study was approved by the ethics board of Linköping University and all participants provided written consent for participation.

**Cell Isolation and Stimulation**

Peripheral blood mononuclear cells (PBMCs) were enriched from healthy donor buffy coats using Lymphoprep (Axis-shield). Human total CD4+ T cells or NT cells then were isolated through magnetic sorting (Miltenyi Biotec). NT cells were cultured with plate-bound anti-CD3 (500 ng/ml) and soluble anti-CD28 (500 ng/ml) in the presence of IL-12 (5 ng/ml), IL-2 (10 ng/ml), and anti-IL4 (5 μg/ml) for Th1 or IL-4 (10 ng/ml), IL-2 (10 ng/ml), anti-IFNγ (5 μg/ml), and anti-IL12 (5 μg/ml) for Th2 conditions. Cells were grown for up to 5 days in RPMI-1640 medium (Life Technologies) supplemented with 2 mM L-glutamine (PAA Laboratories), 10% heat-inactivated fetal calf serum (FCS, PAA Laboratories), and 50 μg/ml gentamicin (Sigma-Aldrich). Cells were re-stimulated with respective condition (see above) after 3 days in culture.

**DNA and RNA Extractions**

Total RNA and genomic DNA were extracted using AllPrep DNA/RNA mini kit (Qiagen). DNA and RNA integrity was determined using a 2100 Bioanalyzer (Agilent Technologies).

**hMeDIP-Seq and Data Analysis**

DNA (1.5 μg) was fragmented by sonication using a Bioruptor (6 × 15 min/30 s on/off, Diagenode), and then it was subjected to end repair, adapter ligation using the NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs). Samples were subjected to hMeDIP as previously described (Nestor et al., 2012). Briefly, DNA was denatured and incubated with 1 μg antibody against 5hmC (ActiveMotif, 39769) overnight at 4°C. Immunoprecipitate and input DNA were prepared using magnetic beads (Dynabeads Protein G, Invitrogen) and amplified using NEBNext Multiplex Oligos for Illumina (New England Biolabs). Samples were separated by electrophoresis on a 2% low-melting-point agarose gel and fragments between 100 and 400 bp were selected. DNA was purified using the QIAquick Gel Extraction Kit (Qiagen) and paired-end reads were sequenced on an Illumina HiSeq 2000 platform at the Beijing Genomics Institute. Reads were mapped to the human genome (hg19) using bowtie2 fast read aligner with the following parameters (bowtie –v 1 – best -S).

Peak calling was performed using model-based analysis of ChIP-seq (MACS) with input samples as controls. NT input was used as input control for both NT and day 1 IP samples. Day 5 input was used as input control for day 5 IP samples. Differential analysis of hMeDIP-seq between conditions was performed using macs2diff algorithm, a sub-program of the MACS software using a standard window size of 200 bp (Zhang et al., 2008). Metagene plots of average 5hmC enrichment were created using the ngs.plot software (Shen et al., 2014). Overlap of 5hmC peaks with different genomic compartments was performed using the intersect function within the bedtools suite of programs (Quinlan and Hall, 2010). Merging of intervals from replicate experiments was performed using the merge function, also within bedtools. Human transcript coordinates were obtained from the refGene database (UCSC Genome Browser, hg19, May 2015; https://genome.ucsc.edu).

**Immuno-dot Blotting**

DNA was denatured and applied to a positively charged nylon membrane under vacuum using a Dot Blot Hybridization Manifold (Harvard Apparatus). The
membrane was washed twice in 2X SSC buffer, air dried, and UV crosslinked. Membranes were incubated with an antibody against 5mC (1:3,000, Active Motif) for 1 hr at 4°C, then washed in TBS-Tween (0.05%), and incubated with a horseradish peroxidase (HRP)-conjugated goat-anti-rabbit antibody (1:10,000, Bio-Rad). Following treatment with enhanced chemiluminescence (ECL) substrate, membranes were scanned on a ChemiDoc MP imaging system (Bio-Rad). To control for loading, membranes were stained with methylene blue. Spot intensities were quantified using ImageJ (NIH).

Gene Expression Microarrays and Analysis
For gene expression microarrays, RNA was labeled and amplified using the Low Input Quick Amp Labeling kit (Agilent Technologies), then hybridized onto SurePrint G3 Human Gene Expression 8x60K v2 microarrays (Agilent Technologies), and scanned using a Surescan High Resolution DNA Microarray Scanner (Agilent Technologies). Raw intensities were exported with Agilent’s Feature Extraction Software. All subsequent analyses were performed using the LIMMA package in the R statistical programming language. Briefly, data were background corrected and quantile normalized; then control probes, probes not expressed (background +10%) in all conditions, and non-annotated probes were removed.

Statistical Analysis
The p values < 0.05 were considered significant and, when stated, p values were adjusted using Benjamini-Hochberg correction.

ACCESSION NUMBERS
The accession numbers for the microarray and next-generation sequencing data reported in this paper are ArrayExpress: E-MTAB-4685, E-MTAB-4686, E-MTAB-4687, E-MTAB-4688, and E-MTAB-4689.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, five figures, and nine tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.05.091.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS
M.S. and B.K. are employees of Genomatix Software. Work in C.E.N.’s lab is supported by grants from the Swedish Research Council and the Åke Wiberg’s Foundation. Work in M.B.’s lab is supported by grants from the Swedish Research Council and Cancerfonden. H.L. was supported by the grant the clinical cooperation group “Nutrigenomics and Type 2 Diabetes” received from the Heilmholt Zentrum München and the Technische Universität München. Work in R.R.M.’s lab is supported by the Biotechnology and Biological Sciences Research Council (BBSRC), CEFIC, and the Medical Research Council (MRC).

Received: December 21, 2015
Revised: March 24, 2016
Accepted: May 22, 2016
Published: June 23, 2016

REFERENCES


Wellcome Trust Case Control Consortium, & Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678.


