Single-Molecule DNA Methylation Reveals Unique Epigenetic Identity Profiles of T Helper Cells

Chloe Goldsmith, Valentin Thevin, Olivier Fesneau, Maria I Matias, Julie Perrault, Ali Hani Abid, Naomi Taylor, Valérie Dardalhon, Julien C Marie, Hector Hernandez-Vargas

HHS Public Access

Author manuscript

J Immunol. Author manuscript; available in PMC 2024 September 15.

Published in final edited form as:

J Immunol. 2024 March 15; 212(6): 1029–1039. doi:10.4049/jimmunol.2300091.

Single molecule DNA methylation reveals unique epigenetic identity profiles of T helper cells

Chloe Goldsmith[*,†], **Valentin Thevin**[*], **Olivier Fesneau**[*], **Maria I. Matias**[‡], **Julie Perrault**[‡], **Ali Hani Abid**[*], **Naomi Taylor**[‡,¶], **Valérie Dardalhon**[‡], **Julien C. Marie**[*], **Hector Hernandez-Vargas**[*,||]

[*]Tumor Escape Resistance and Immunity Department, Cancer Research Center of Lyon (CRCL), INSERM U1052, CNRS UMR 5286, Centre Léon Bérard (CLB) and University of Lyon 1, Lyon, France. Labellisee Ligue Nationale Contre le Cancer.

[†]Current address: Exercise Epigenetics, University of Canberra Research Institute for Sport and Exercise (UCRISE), ACT, Australia.

[‡]Institut de Génétique Moléculaire de Montpellier, Univ Montpellier, CNRS, Montpellier, France.

[¶]Pediatric Oncology Branch, NCI, CCR, NIH, Bethesda, MD, USA.

[||]Genomics Consulting, 69500 Bron, France.

Abstract

Both identity and plasticity of CD4 T helper (Th) cells are regulated in part by epigenetic mechanisms. However, a method that reliably and readily profiles DNA base modifications is still needed to finely study Th cell differentiation. Cytosine methylation in CpG context (5mCpG) and cytosine hydroxymethylation (5hmCpG) are DNA modifications that identify stable cell phenotypes, but their potential to characterize intermediate cell transitions has not yet been evaluated. To assess transition states in Th cells, we developed a method to profile Th cell identity using Cas9-targeted single molecule nanopore sequencing. Targeting as few as 10 selected genomic loci, we were able to distinguish major *in vitro* polarized murine T cell subtypes as well as intermediate phenotypes by their native DNA 5mCpG patterns. Moreover, by using off-target sequences we were able to infer transcription factor activities relevant to each cell subtype. Detection of 5mCpG and 5hmCpG was validated on intestinal Th17 cells escaping TGFb control, using single-molecule adaptive sampling. 21 differentially methylated regions mapping to the 10-gene panel were identified in pathogenic Th17 cells relative to their non-pathogenic

Author Contributions

CG conducted all sequencing experiments.
MM and JP performed all in vitro differentiation experiments.
JM, OF and VT participated in study design.
CG, AH and HH performed bioinformatic analyses
CG and HH wrote the manuscript.
VD, NT and HH supervised the work.
HH conceived the study.
All authors discussed the results and manuscript text.

Competing interests

CG and HH have received travel and accommodation support to attend conferences for Oxford Nanopore Technology. The authors declare no additional competing interests.

counterpart. Hence, our data highlight the potential to exploit native DNA methylation profiling to study physiological and pathological transition states of Th cells.

Keywords

DNA methylation; 5mC; DNA hydroxymethylation; 5hmC; Th cells; long reads; nanopore sequencing; immune cells; adaptive sampling

Introduction

In general, adaptive immunity in both protective and autoimmune settings is orchestrated by the emergence of polarized effector CD4 T helper (Th) lymphocyte subsets of specialized function. Naive CD4 Th0 cells under the influence of cytokines and TCR signaling can give rise to different effector cells including Th1,Th2, Th17, and T regulatory (Treg) cells (1, 2). Furthermore, plasticity is an important property of Th cells, whereby they adapt their functions in response to a changing environment forming a continuum of polarized phenotypes (3). Hence, cells with intermediate Th1/Th17 and Th17/Treg phenotypes have been described in both mice and humans in association with autoimmune diseases and protective functions, respectively (4).

Both identity and plasticity of Th cells are regulated in part by epigenetic mechanisms (2, 5). Epigenetics refers to the heritable layer of information on top of a DNA sequence that regulates gene expression and makes each cell type unique. Therefore, mapping epigenetic modifications represents an ideal way to discriminate cell types by their ‘epigenotype’ (6, 7). DNA methylation (5mC) is the most stable and most widely studied epigenetic mark and is involved in gene silencing (8). DNA hydroxymethylation (5hmC) is relatively understudied, but recently has been linked to gene activity during differentiation (9, 10). Because of their relative stability (11) and cell-type specificity (9), mapping 5mC/5hmC landscapes has the power to improve our understanding of Th cell identity and plasticity, and aid in the development of novel biomarkers for these cells.

Despite the evidence for 5mC/5hmC as a marker of cellular identity, in particular in Th cells, technical limitations have prevented this knowledge from reaching clinical use. Simultaneous detection of these cytosine modifications in CpG context (5mCpG and 5hmCpG) has been made recently available by the next generation of Nanopore long-read sequencers, which are able to provide such information in native DNA. Here, we take advantage of Nanopore sequencing to simultaneously profile 5mCpG and 5hmCpG in native DNA on naive Th cells polarized under Th0, Th1, Th2, Th17 and Treg conditions. In addition, given that Transforming Growth Factor beta (TGFb) is known to prevent pathogenic Th17 cell differentiation (12), we performed targeted sequencing on both, *in vitro* and *in vivo* generated Th17 cells escaping TGFb control. This study opens new avenues towards the implementation of 5mC/5hmC in clinical immune profiling and identification of pathogenic immune cell subsets.

Materials and methods

Mice

FoxP3-GFP reporter male mice (13) between the ages of 6–12 weeks of age were used for in vitro Th polarization. Il-17a-CRE mice (14) were bred with mice bearing floxed alleles of TgfbR2 (15). These so called TGFbR-KO mice and their TGFbR-WT counterpart (littermates without floxed alleles) were then crossed with Rosa-26-stopFlx/Flx-yfp mice to enable tracking by monitoring yellow fluorescence protein (YFP) expression, which represents the cells that either are expressing or expressed Il17a.

All animals were housed in ventilated racks in pathogen-free facilities at the Institut de Génétique Moléculaire de Montpellier (Montpellier, France) and P-PAC (Lyon, France). Animal care and experiments were approved by the local and national animal facility institutional review boards in accordance with French national and ARRIVE guidelines (CLB 2017–017).

Cell isolation

Small intestine colon and mesenteric lymph (mLN) nodes from TGFbR-WT and TGFbR-KO mice were dissected. Fat was removed and intestines were longitudinally opened and washed in PBS 1X (Gibco). Intestines were cut into small pieces and incubated with 5mM EDTA, 1mM DTT (D0632, Sigma Aldrich) at 37°C. Epithelial cells were then separated from intra-epithelial lymphocytes with a 44% / 67% percoll gradient (P1644, Sigma Aldrich) run for 20 minutes at 1300 x G. Tissues were then digested in RPMI medium (Gibco) containing 20% fetal bovine serum (Gibco), DNAse I (Roche) at 100µg/ml and collagenase from Clostridium Histolyticum (C2674, Sigma Aldrich) at 1mg/ml. Intestinal Lamina propria lymphocytes were then separated on a 44% / 67% percoll gradient run for 20 minutes at 1300 x G. mLN cells were ground between wire mesh. CD4+ T cells were enriched using the CD4+ T cell Isolation Kit mouse (Miltenyi Biotec.), and then labeled with CD4 PE (GK1.5, eBioscience) TCRb APC (H57–597, BD Bioscience), CD45 APC-Cy7 (30-F11, BD biosciences) and DAPI (Eurobio scientific). Cells were purified on ARIA II (BD Biosciences). For in vitro polarization CD4+ T cells were purified from Foxp3-GFP mice using the MACS CD4+ T cell negative selection kit (Miltenyi Biotec) and naïve CD4+ T cells from mice were then sorted on the basis of a CD4+CD8-CD62L+CD44−GFP-CD25- expression profile on a FACSAria flow cytometer (BD Biosciences).

In vitro Th polarization

Purified naïve CD4+ T cells from male Foxp3-GFP mice were activated under different polarizing conditions during 4 days: Th0 (neutral), Th1, Th2, Th17, Treg as well as Th17s differentiated without TGFb1 (R&D Systems, Cat#240-B) (Th17-noTGFb) representing an intermediate cell phenotype (Supplemental Tables 1 and 2). Activation was performed using plate-bound anti-CD3 (clone 145–2C11, 1 µg/ml [BioXcell, Cat# BE0001–1]) and anti-CD28 (clone PV-1, 1 µg/ml [BioXcell, Cat# BE0015–5]) monoclonal antibodies in RPMI 1640 medium (Life Technologies) supplemented with 10% FCS, 1% penicillin/streptomycin (Gibco-Life technologies) and beta-mercaptoethanol (50 µM). Cells were split 3 days later with medium supplemented with rhIL-2 (100U/mL). Cells were maintained in

a standard tissue culture incubator containing atmospheric O2 and 5% CO2. 5MCpG raw data from Tregs included in this study was previously published by us under the exact same experimental conditions (5). These data were reanalyzed to compare an additional CD4 Th subset using updated basecalling algorithms and analysis pipelines.

Flow cytometry and Cytometric Bead Array (CBA)

Immunophenotyping of cells was performed with fluorochrome-conjugated antibodies, and intracellular staining was performed after fixation and permeabilization (ThermoFisher). Cells were labeled with fixable viability dye prior to fixation, followed by at least 3 hours of intracellular staining with the following anti-transcription factor antibodies: Foxp3-PeCy7, T-Bet-PE, Rorγt-BV650 and Gata3-APC (Supplemental Table 1). Cytokine production (IFNg, IL-4, IL-17A, IL-10) was also assessed by Cytometric Bead Array (CBA) Kit (BD Biosciences) on supernatants collected at day 3 of polarization. Data analysis was performed using FlowJo (Tree Star software) and FCAPArray Software (CBA analysis).

DNA extraction

DNA was extracted from 1×10^6 CD4 Th cells/replicate using the Qiagen DNA mini kit according to manufacturer's instructions (Qiagen Hilden, Germany). DNA was extracted from 10ng of mouse brain tissue as previously described (16) using NEB Monarch High molecular weight DNA extraction kit according to manufacturer's instructions (NEB Cat #T3050). Viable CD45+CD4+YFP+ cells were sorted from the small intestine, lamina propria, Peyer's patches and mesenteric lymph nodes of 6–9 months old mice using a FACS ARIA II (BD Biosciences). Beckman GenFind V3 DNA extraction kit (Beckman Coulter) was used to isolate DNA from 120,000 TGFbR-WT and 180,000 TGFbR-KO sorted cells, following manufacturer's instructions.

Library prep and whole genome Nanopore sequencing

DNA was sheared using g-tubes (Covaris, USA) to ~10Kb before library preparation with ligation sequencing kit (ONT Cat #SQK-LSK109) followed by whole genome sequencing on Nanopore MinION flow cells with pore 9.4.1 chemistry for 72h. Raw read quality was determined with pycoQC (17).

Cas9 enrichment, library prep and targeted Nanopore sequencing

A double cutting approach was used whereby 2x S. pyogenes Cas9 Alt-R™ trans-activating crispr RNAs (crRNAs) were designed for both upstream and downstream of each target loci as previously described (5, 18) (crRNA details and sequences provided in Supplemental Table 3). crRNAs were pooled in equimolar ratios (100 μM) and annealed to S. pyogenes Cas9 Alt-R™ tracrRNA (100 μM) (Integrated DNA Technologies, Iowa United States). crRNA•tracrRNA pool (10 μM) was incubated with Cut smart buffer (New England Biosciences Cat #B7204) and Alt-R® S. pyogenes HiFi Cas9 nuclease V3 (62 μM) forming crRNA-tracrRNA-Cas9 ribonucleoprotein complexes (RNPs). 3000 ng of high molecular weight DNA was prepared by blocking available DNA ends with calf intestinal phosphatase (NEB Cat #M0525). DNA and dA-tails on all available DNA ends were cleaved by RNPs and Taq polymerase (New England Biosciences Cat #M0273) and dATP (New England

Biosciences Cat #N0440), activating the Cas9 cut sites for ligation. Nanopore Native Barcodes (Oxford Nanopore Technology Cat #EXP-NBD104) were ligated to available DNA ends with NEB Blunt/TA Ligase Master Mix (New England Biosciences Cat #M0367) and up to 4 samples were pooled before ligation of Nanopore adapters with NEBNext® Quick T4 DNA Ligase (New England Biosciences Cat #E6057) and loading onto a primed MinION flow cells (pore chemistry 9.4.1) and sequencing for 72h. Raw read quality was determined with pycoQC (17).

For adaptive sequencing we used the 'UNCALLED' methodology in targeted reference genes (19). To this end, a reference gene panel was developed based on known targets of TGFb (20). Gene sequences were retrieved from the mouse mm10 assembly, including a 10KB flanking region up and downstream of each region of interest, and masking repeated elements. This method did not require Cas9-based library preparation. Instead, during sequencing, UNCALLED is able to rapidly match streaming of nanopore current signals to the reference sequence, enriching in this way for our selected 139 targets. Full reference list and UNCALLED loci are listed in Supplemental Table 4.

Detection of 5mC and 5hmC from Nanopore signal data

Raw fast5 files were converted into pod5 format and basecalled with 'dorado' (version 0.4.2, ONT), allowing the simultaneous delineation of 5mCpG and 5hmCpG using a high accuracy model ('dna_r9.4.1_e8_hac@v3.3', ONT) for the detection of modified bases (18). Dorado was also used for simultaneous alignment to the mm10 assembly and demultiplexing, followed by sorting and indexing with 'samtools' (version 1.13). 'Modkit' (version 0.2.2) was used for piling of methylation data into bed files.

Detection of differential methylation

Data processing and statistical analyses were performed using R/Bioconductor (R version 4.3.1). Bed files containing modified cytosine information were transformed into a BSseq object for differential methylation analysis with dispersion shrinkage for sequencing data (DSS) as previously described (5, 21). DSS tests for differential methylation at single CpG-sites were evaluated using a Wald test on the coefficients of a beta-binomial regression of count data with an 'arcsine' link function. Differentially methylated regions (DMRs) were defined as those loci with at least 3 CpG sites within a distance of less than 50 bp, and with changes in > 50% of all CpG sites exhibiting p value < 0.05. DMRs were plotted using the plotDMRs function of the dmrseq R package (22, 23). Off-targets were analyzed for transcription factor (TF) activity using the cistrome information associated with the MIRA R package (24). After extracting all genomic binding regions corresponding to selected TFs, MIRA functions were used for aggregation of DNA methylation data, estimation of TF scores and visualization of inferred TF activity. Only TFs described as high confidence were used, with a coverage of at least 100x and at least 10,000 annotated targets.

Results

Targeted Nanopore sequencing of in vitro polarized CD4 Th cells achieves high coverage of key loci

To study the relationship between DNA methylation and classic Th cell identity, we polarized naïve cells towards various Th fates in vitro. Importantly, this approach does not lead to a terminal differentiation state but rather a polarization towards different fates (25). Naïve cells were isolated by FACS-sorting and activated under polarizing conditions towards the Th0, Th1, Th2, Th17 and Treg phenotypes using cytokine cocktails. To test the sensitivity of this approach to detect potential pathogenic Th cell phenotypes, naïve T cells were also polarized under Th17 conditions in the absence of TGFb, referred to as 'Th17-noTGFb' cells (Fig. 1A).

Specific transcription factor staining (Fig. 1B-C) and cytokine production (Supplemental Fig. 1A) associated to Th cells were performed to estimate the efficacy of polarization for each condition. Expression patterns of transcription factors T-bet, Gata3 and Foxp3 appeared selectively upregulated in Th1, Th2 and Treg polarizing cells, respectively (Fig. 1B-C) and correlated with the cytokine signature detected (Ifng and Il4 for Th1 and Th2 conditions, respectively; Supplemental Fig. 1). Expression of Rorgt was consistently upregulated in Th17 and Th17-noTGFb polarized cells (Fig. 1B-C). MFI of Rorgt increased by 1.8 fold in Th17 and 1.7 fold in Th17-noTGFb as compared to control conditions. Furthermore, only cells polarized towards a Th17 fate produce significant amounts of Il17 and did not produce Ifng or Il4 at a detectable level (Fig. 1C). Taken together, these data indicate an efficient polarization of Th17 and Th17-noTGFb conditions.

To characterize the DNA methylation landscape of polarized Th cells, we developed a Cas9 guided capture method coupled with Nanopore sequencing (Fig. 1A). We manually selected key genes that could potentially be used to classify different Th subsets. This panel included DNA regions whereby differential epigenetic modifications in Th cell subsets have been previously reported (26–28). In general, these genes code for major transcription factors (TF) and cytokines specific to Th cell subsets, namely Il17a, Maf, Rorgt/Rorc and Zfp362 (28) (Th17); Foxp3, Il10 and Itgb8 (Treg); Gata3, Il4 and Il10 (Th2); Tbx21/T-bet and Ifng (Th1). Cas9 guided enrichment captures gene length regions prior to single molecule native DNA sequencing with Nanopore. This involved the design of 2 sgRNA's upstream and downstream of each gene of interest employing a double cutting approach. It was then followed by synthesis of Cas9-guided Ribonucleotide proteins (RNPs) to enrich for targets prior to Nanopore sequencing (Fig. 1D) as previously described (5, 7, 21). Up to 4 samples from each polarization assay (i.e. Th0, Th1, Th2, Th17, Th17-noTGFb and Tregs) were barcoded and multiplexed on each MinION flow cell, and profiled for 5mCpG and 5hmCpG landscape of whole gene length single DNA molecules. In line with previous work on other cell types, the coverage of 10x was sufficient to detect a 25% change in DNA methylation (7) and was therefore set as the minimum target coverage for this study. The typical sequencing yield achieved for each of these Nanopore runs was ~1GB, which is in accordance with the literature (7, 18). Importantly, we exceeded the minimum coverage threshold of 10x for target regions (average coverage 45x, Fig. 1D and Supplemental

Fig. 1B) and obtained full gene length reads (N50 = 21KB) on all targets (Fig. 1E and Supplemental Fig. 1C); this was sufficient for further processing to detect differentially methylated regions.

These data highlight the efficacy of Cas9 targeted sequencing for capturing important regions for Th cells.

Classification of Th cell subsets using DNA methylation (5mCpG)

To basecall raw reads and detect modified bases, we used the latest high performing nanopore tool, Dorado (https://github.com/nanoporetech/dorado) combined with high accuracy models to call DNA modifications. Dorado works by extracting modified base calls from raw nanopore reads and anchoring the information-rich basecalling neural network output to a reference genome (mm10). Global patterns of methylation were compared between different Nanopore sequencing runs and no differences were detected (Fig. 1F and 1G). In all sequencing runs 5mCpG displayed a typical bimodal distribution with peaks close to zero and 100% methylation, and a variable fraction of CpG sites with intermediate levels of methylation (Fig. 1F). In contrast, 5hmCpG was characterized by very low levels in all sequencing runs, with a much lower fraction of detectable methylated sites above 10% (Fig. 1G).

Differentially methylated loci (DMLs) between subsets were defined as individual CpG sites with significant change (p value < 0.05) of at least 10% in any direction. The number of distinct DMLs identified in Th1, Th2, Th17, and Treg cells when comparing to Th0 cells, were 305, 359, 428, and 490, respectively. Furthermore, a comparison of Th17 cells polarized in presence or absence of TGFb (i.e. Th17-noTGFb cells) identified 378 DMLs, highlighting a potential effect of TGFb alone on modulating DNA methylation during T cell polarization.

To obtain a more robust measure of 5mCpG variation, we searched for differentially methylated regions (DMRs), defined as genomic ranges of at least 50 bp and a minimum of 3 CpG sites, where at least half of the CpG sites are significantly different ($p < 0.05$ and 10% methylation change) between two cell subsets. This analysis resulted in 20 DMRs that distinguished Th0 cells from the different polarized subsets (Supplemental Table 5). Two additional DMRs were identified in Th17-noTGFb cells relative to Th0 cells and Th17 cells polarized in the presence of TGFb. The length of these DMRs averaged 214 bp, and ranged from an off-target 53 bp-region in the *Tnk2* locus, to a large 906 bp DMR in the *Il4* locus displaying 60% demethylation in Th2 cells relative to Th0 cells. In terms of CpG content, they ranged from 4 to 15 sites and an average of 6 CpG sites per DMR. In line with the selection of the targeted DNA regions, DMRs were largely negative compared to Th0 cells, highlighting that in these identity genes, 5mCpG decreased during T cell differentiation in correlation with increased gene expression. Significantly, hierarchical clustering of the resulting 22 DMRs was sufficient to cluster Th cells into their respective subsets (Fig. 2A). Interestingly, Th2 cells were the most divergent of the different cell subsets in this clustering analysis (Fig. 2A).

Plots of each DMR (Fig. 2B-F) illustrate the specificity of methylation profiles for each cell type. Indeed, DMRs upstream of the *Foxp3* promoter and in a CpG island (CGI) shore on *Tbx21* were unique to Treg and Th1 cells, respectively (Fig. 2B and 2C). Eight DMRs distinguished Th2 cells, three of them in the *Il4* gene body and its promoter (Fig. 2D, upper panel) and the others in *Il10*, *Gata3* and close to the *Foxp3* locus. Six DMRs were associated with Th17 polarization, including one in a distal enhancer close to *Zfp362* and two in the body of the *Rorc* gene (Fig. 2E). The same *Zfp362* enhancer distinguished Th17 cells polarized in presence or absence of TGFb (Fig. 2F), with *Zfp362* remaining hypomethylated in Th17_noTGFb cells at levels close to Th0 cells.

These data highlight the power of characterizing DNA methylation profiles in a small subset of genes as means to determining Th cell identity. They also suggest that 5mCpG profiles can be used to evaluate the specific impact of a cytokine on Th differentiation.

DNA hydroxymethylation (5hmCpG) does not add discriminatory capacity to 5mCpG profiling

Though 5hmCpG is generally associated with increased gene expression when present in coding regions (21), the relationship between 5mCpG and 5hmCpG in CD4 Th gene regulation is poorly described. This is of importance since 5mC and 5hmC have been shown as key players in tissue differentiation and cellular plasticity (9). Therefore, we took advantage of Nanopore to simultaneously detect 5mCpG and 5hmCpG on the same native Nanopore reads, as described in Methods. While the detection of 5mCpG with Nanopore has been validated, the detection of 5hmCpG has not. Given that tissue derived from the central nervous system (CNS) is known for being high in DNA hydroxymethylation (29, 30), we used data obtained from DNA extracted from the brain of P13 OF1 mice as a positive control for 5hmCpG. Using the same pipeline described above for calling modified bases, a small fraction of CpG sites (3%) displayed 5hmCpG of at least 10% in brain tissue. This percentage increased to 4.2% when aggregating the 5hmCpG data corresponding to forebrain enhancer regions. This was significantly higher than the 3.3% average 5hmCpG identified at control adipose enhancer regions ($p < 0.05$, Supplemental Fig. 1D and 1E). Although this is far from a rigorous validation of nanopore-based 5hmCpG accuracy, we proceeded to determine the 5hmCpG landscape of CD4 Th cell subsets in an exploratory manner.

No differentially hydroxymethylated regions (5hmCpG DMRs) were identified with the same criteria used earlier for 5mCpG. Despite of this, 42 regions contained significant CpG sites in discrete regions, regardless of their magnitude of methylation change (Supplemental Table 6). Together, these regions do not discriminate cell subtypes, as opposed to 5mCpG (Fig. 3A and 3B). Again, two of the three Th2-polarized samples are clustered apart, in particular due to few loci mapping to *Il4, Il10, Zfp263* and *Gata3* (Fig. 3A). Of note, most of these regions displayed higher 5hmCpG levels in Th2 cells, in contrast to their lower 5mCpG levels described above.

Based on these results, 5hmCpG does not add discriminatory capacity to 5mCpG at the loci and conditions of this analysis.

DNA methylation in transcription factor binding sites reveals their activity in Th cells

In addition to targeted regions, Nanopore sequencing captures around 1GB of 'off' target reads in each sequencing run. To strengthen the identification of Th cell subsets using DNA methylation, we determined the DNA methylation status of TF-binding sites in 'on' and 'off' target reads with Bioconductor package MIRA (Methylation-based inference of regulation activity) (24). MIRA aggregates the DNA methylation values in all known binding sites for chosen TFs and uses this profile to produce a score. Such score infers the level of regulatory activity on the basis of the shape of the DNA methylation profile, based on the observation that DNA methylation tends to be lower in regions where TFs are bound (31). In a first step, TF's were chosen matching the known transcriptome orchestrators of Th cell subtypes Tbx21 and Stat1 (Th1); Gata3 (Th2), Stat3 and Runx3 (Th17), and Foxp3 (Treg) (32, 33). Bcl6 is a TF for T follicular helper cells (33) and was used as a negative control for target Th cells. We aggregated 5mCpG located at known binding sites for the chosen TFs, representing thousands of regions (Fig. 4, left panels). In general, we observed a hypomethylation dip for all TFs, indicating regulatory activity. In addition to plotting MIRA profiles, we determined the MIRA scores for each cell subset presented as boxplots next to the aggregated methylation plots (Fig. 4). Three of the selected TFs fell below statistical significance (i.e. FOXP3, GATA3 and RUNX3), with Tregs displaying the highest MIRA scores for FOXP3 binding sites, Th2, Tregs and Th17 displaying high scores for GATA3 binding sites, and Th17 cells having the highest score for RUNX3.

To provide an unbiased analysis of off-target sequences, we next performed the same analysis for all high-confidence TFs available in the MIRA database. Out of 877 TFs, only 181 are classified as high confidence by MIRA's own classification. After keeping only those TFs with more than 10k targets across the genome and at least 100x coverage in our data, we inferred the scores for 87 TFs. Only two TFs were significantly different after p value correction (FDR < 0.05), IKZF1 and PAX5 (Figure 4). In both cases, Th17 cells displayed the highest MIRA scores.

Together, these data show that the methylation status of TF binding sites can be used as an indicator of the activity of Th cell subsets, further strengthening the relationship between 5mC and Th cell identity.

Single molecule DNA methylation in Th17 cells *in vivo*

Our results above suggest that TGFb is partially involved in shaping the methylome landscape during polarization of Th17 cells. In order to validate our results *in vivo*, we used Th17 cells present in the gut of Il17aCRE; Tgfbr2 fl/fl Rosa-26 stop fl/fl- yfp (TGFbR-KO) mice and Il17aCRE; Rosa-26 stop fl/fl- yfp (TGFbR-WT) mice. These cells escape TGFb signaling control and can be purified from the tissues based on YFP expression. Their methylation profiles were analyzed through a modified method of targeted sequencing, where single molecules are selected during nanopore sequencing with coupled real time basecalling (see details in Methods). This so-called adaptive sampling was used to extend the selection of genomic regions from the original panel of 10 genes to 139 annotated genes, selected because of their known role in TGFb signaling (20). We compared 5mCpG in TGFbR-KO Th17 cells to their WT counterpart. A total of 330 DMRs distinguished the two

conditions, with 21 of them within the original 10-gene panel (Fig. 5A and Supplemental Table 7). In line with the pathogenic potential linked to the absence of TGFb control on Th17 cells (12), *Tbx21,* encoding for T-bet, was the top most significant of these 10 genes, with 8 DMRs falling within its locus and 7 of them showing hypomethylation in TGFbR-KO cells (Fig. 5B). Moreover, two of these DMRs displayed differences in methylation of more than 60% (hypomethylated in TGFbR-KO cells relative to TGFbR-WT cells). Similarly, in line with a epigenetic signature of *in vivo* generated Th17 cells (28), three DMRs were identified in the *Zfp362* locus, including a large hypomethylated region upstream of the *Zfp362* promoter that overlaps with the DMR identified in *in vitro* differentiated Th17 cells in the absence of TGFb (Fig. 5B). One of these DMRs matched a distal enhancer region and was at least 60% less methylated in TGFbR-KO cells relative to TGFbR-WT cells.

5hmCpG was also studied in the Th17 cells from the mouse gut. Overall, 5hmCpG was detectable in a higher proportion of CpG sites and higher levels than *in vitro* polarized cells, although no DMRs were detected in any of the genes of the original 10-gene panel. When expanding the search to the 139 targeted regions, three 5hmCpG DMRs were identified in two different genes, *Zeb2* and *Pparg*. One of these, a 173 bp region in the *Zeb2* locus, spanned 8 CpG sites with 5hmCpG levels that increased from an average of 3% in TGFbR-WT cells to 38% in TGFbR-KO cells. The same region was identified as a 5mCpG DMR hypomethylated in TGFbR-KO cells (Supplemental Table 7). Thus, 5hmCpG can indeed be detected in Th17 cells at discrete loci *in vivo*.

This set of data highlights the potential to exploit native DNA methylation profiling to study physiological and pathological features of Th17 cells present in tissues.

Discussion

DNA cytosine modifications are known to be cell-type specific(1). However the role of these modifications in Th cell identity is not well understood. The present study addressed the relationship between DNA methylation and Th cell identity, characterizing 5mCpG and 5hmCpG of murine CD4 Th cell subsets (Th0, Th1, Th2, Th17, Th17-noTGFb and Tregs) by long read single molecule native DNA sequencing. We propose a novel approach able to identify differentially methylated regions expected for each Th cell subset. Using this method, our data revealed that 5mCpG profiles from different loci provide a map specific to each in vitro-polarized Th cell subtype. Moreover, we confirmed the efficiency of nanopore targeted sequencing to discriminate pathogenic and non pathogenic Th17 cells *in vivo*.

We developed a Cas9 targeted sequencing to enrich important loci for Th cell identity. This method obtained high coverage of target genes compared to short read sequencing approaches, as well as whole gene length reads. Cas9 enrichment is an innovative technique that has been adopted to answer a variety of biological questions ranging from studying epigenetic heterogeneity in viruses (7), important loci for cancer (18) and neuromuscular disorders (34). Regardless of the question, Cas9 targeted enrichment coupled with Nanopore sequencing obtains anywhere from 0.5–2GB of data and 10 – 1000s of on target reads. In the present study, we observed variability between coverage of different target loci achieving the highest coverage of *Il4*, while the lowest with *Itgb8*. This coverage disparity is likely

due to differences in size of target regions. *Il4* is a smaller gene and thus to obtain all regulatory information nearby we aimed to enrich in ~7Kb, while *Itgb8* is a larger gene and thus required enrichment of >100Kb. Shorter fragments elicit a higher coverage due to the nature of Nanopore's preferentially sequencing shorter reads. Nevertheless, the on-target reads obtained in the present study fall well within that of the literature (18, 34).

Our data shows that 5mCpG in a selection of genes can accurately classify polarized Th cell subsets. Importantly, capturing long reads with high coverage of important loci revealed that DMRs were not only specific for a single subset; instead, some regions exhibited a range of methylation profiles. Other techniques relying on sodium bisulfite conversion are not able to distinguish 5mC from other modified bases such as 5hmC, which could lead to erroneous conclusions regarding the relationship between different modified bases and their role in gene regulation. Therefore, we focused our discussion on studies that consider both 5mC and 5hmC. Several studies have investigated the DNA methylation landscape of certain subsets (Th1 and Th2) using BS-seq and hydroxymethylation with immunoprecipitation techniques (hMeDIP) (35). The authors highlighted, Th1 and Th2 cells could be distinguished by 5mC and 5hmC changes in important identity genes. We furthered this work by mapping 5mC and 5hmC in important loci for these and additional Th cell subsets (Th17, Th17-noTGFb and Tregs) and identified unique 5mCpG patterns for each subset, and thus was able classify them by their epigenetic landscapes. Importantly, by characterizing methylation profiles in our identity gene panel, we were able to discriminate the unique Th17-noTGFb cells indicating that an epigenetic approach has the potential to identify additional pathogenic and intermediate Th cell subsets. This is line with other studies showing that TGFb in cultured cells may influence methylation profiles through cell division (36). Moreover, our own findings confirm that large differences in 5mCpG can also be identified in vivo in cells that escape TGFb control. Emerging evidence suggests metabolomics and epigenetics work in concert to alter Th cell identity (1).

DNA hydroxymethylation is an emerging marker of cellular identity and has been studied largely in the context of cellular differentiation (10, 37). For detection of multiple modified bases (5mC, 5hmC etc), most techniques require samples to be split, and different modified bases to be detected separately. In the present study, we took advantage of Nanopore sequencing's ability to determine the 5mCpG and 5hmCpG simultaneously. We found that 5hmCpG across a selection of genes was able to classify T cell subsets to a certain extent, and was largely concentrated in 'active' genes. Our findings are in accordance with recent studies also highlighting that 5hmCpG in T cells is concentrated in the body of active genes (27, 35). Such studies indicated that terminally differentiated Th1 and Th2 cells have very little 5hmCpG (27) compared to CD4+, CD8+ or double-positive T cells (38). Our results obtained straight from *in vivo* differentiated gut Th17 cells suggest that the selected targets are able to discriminate between non-pathological and pathological Th17 cells. Of note, compared to the findings in *in vitro* polarized T cells, the differences observed *in vivo* were stronger in number and magnitude. Not only the differences in 5mCpG were larger but we were also able to observe higher basal levels of 5hmCpG and differences above 30% between conditions. For 5hmCpG in particular, our findings are in line with those suggesting that *in vitro* proliferation is associated with a passive loss of 5hmCpG during adaptation to culture conditions (40).

In addition to targeted DNA methylation analyses, we show here that off-target data can be used to effectively infer TF activity. As T cell identity is defined by the synchronized activity of TFs, we illustrate how this can be exploited to complement locus-specific results without additional extraction or sequencing workload. We foresee that a combination of both on-target and off-target data will be a more robust marker of Th cell identity.

An important consideration for interpretation of the present study; our samples were not purified after in vitro polarization and are known to be heterogeneous (cytokine treatment is not 100% efficient). However, we expect stronger discriminatory capacity in purified cell types. In addition, we selected important loci for Th cell subsets through literature review, which relied on known well annotated regions since there has yet to be a long read whole genome sequencing study of T cell subsets. Although we used analysis of off-target loci to support our main findings and develop an unbiased profile, there are potentially additional genomic regions with better discriminatory capacity that are yet to be identified.

In our study we used two different methods for targeted sequencing, CRISPR-Cas9 selection and UNCALLED adaptive sampling. The major difference between these two strategies is the moment where genomic selection occurs, either during library preparation in the case of CRISPR or during the sequencing process for UNCALLED. While CRISPR-based libraries pass randomly through the nanopores in a given flowcell, adaptive sampling relies on real-time electronic selection of the molecules of interest. More precisely, UNCALLED converts stretches of raw signal into k-mers that are used as a query for an index search against the target database of genes of interest. These methods do not affect the calling of canonical and modified bases, and the downstream methods of analyses are indeed very similar. However, UNCALLED offers the advantage of a larger selection of targets (more than 100 in our hands) and a simplified library preparation. Moreover, other methods of adaptive sampling are currently available (41)(Payne et al, 2021) or under development and will likely improve the coverage and efficiency of real-time sampling over pre-sequencing target selection.

In conclusion, by mapping modified DNA bases (5mCpG and 5hmCpG) at single molecule resolution we show that specific epigenetic footprints are able to distinguish some of the major groups of *in vitro* polarized T cells (i.e. Th0, Th1, Th2, Tregs and Th17s). Moreover, we show that intermediary phenotypes associated with dysfunctional immune responses can also be identified by targeted 5mCpG/5hmCpG profiling *in vitro* (Th17-noTGFb cells), and *in vivo* (TGFbR-KO Th17 cells).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The expert assistance of H Tarayre V, S. Rodriguez, and E Guillemot is acknowledged. We thank Profs G. Stockinger, and A.Y Rudensky for providing with Il17-CRE and Foxp3-GFP mice respectively. The help of the core facilities, of the flow cytometry facility, and mouse facility P-PAC of the CRCL is acknowledged.

Funding

This work was supported by the Institut Nationale de la Sante et de la Recherche Medicale (INSERM) Plan Cancer AAP "Single Cell", édition 2019, No. 19CS144-00 (project: Molecular and cellular map all along anorectal cancer development); the Fonds Amgen France pour la Science et l'Humain 2021 (project: Establishing a clinically committed pipeline to identify T cell and tumor epigenotypes associated with cancer and its response to immunotherapy); L'Institut Convergence Plascan (Grant Number ANR-17-CONV-0002); Le LYriCAN+ (INCa-DGOS-INSERM-ITMO cancer_18003); and LabEx DEVweCAN (ANR-10-LABX-0061). CG was supported by the Fondation MSD Avenir, programme ERICAN 2020 (project: Deciphering the epigenetic mechanisms responsible of the Th17 plasticity associated with tumor development).

Data availability

All raw and processed sequencing data are publicly available at the Gene Expression Omnibus (GEO) repository, under accession numbers GSE248540 and GSE190409 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi). Bioinformatics code necessary to reproduce all analyses is publicly available at GitHub (https://github.com/hernandezvargash/Tcell.5mC.ID).

References

1. Goldsmith CD, Donovan T, Vlahovich N, and Pyne DB. 2021. Unlocking the Role of Exercise on CD4+ T Cell Plasticity. Front Immunol 12: 729366.
2. Locksley RM 2009. Nine lives: plasticity among T helper cell subsets. J Exp Med 206: 1643–1646. [PubMed: 19635860]
3. DuPage M, and Bluestone JA. 2016. Harnessing the plasticity of CD4(+) T cells to treat immune-mediated disease. Nat Rev Immunol 16: 149–163. [PubMed: 26875830]
4. Lochner M, Wang Z, and Sparwasser T. 2015. The Special Relationship in the Development and Function of T Helper 17 and Regulatory T Cells. Prog Mol Biol Transl Sci 136: 99–129. [PubMed: 26615094]
5. Matias MI, Yong CS, Foroushani A, Goldsmith C, Mongellaz C, Sezgin E, Levental KR, Talebi A, Perrault J, Rivière A, Dehairs J, Delos O, Bertand-Michel J, Portais J-C, Wong M, Marie JC, Kelekar A, Kinet S, Zimmermann VS, Levental I, Yvan-Charvet L, Swinnen JV, Muljo SA, Hernandez-Vargas H, Tardito S, Taylor N, and Dardalhon V. 2021. Regulatory T cell differentiation is controlled by αKG-induced alterations in mitochondrial metabolism and lipid homeostasis. Cell Rep 37: 109911.
6. Dozmorov MG, Coit P, Maksimowicz-McKinnon K, and Sawalha AH. 2017. Age-associated DNA methylation changes in naive CD4+ T cells suggest an evolving autoimmune epigenotype in aging T cells. Epigenomics 9: 429–445. [PubMed: 28322571]
7. Goldsmith C, Cohen D, Dubois A, Martinez MG, Petitjean K, Corlu A, Testoni B, Hernandez-Vargas H, and Chemin I. 2021. Cas9-targeted nanopore sequencing reveals epigenetic heterogeneity after de novo assembly of native full-length hepatitis B virus genomes. Microb Genom 7: 000507.
8. Calle-Fabregat C. de la, Morante-Palacios O, and Ballestar E. 2020. Understanding the Relevance of DNA Methylation Changes in Immune Differentiation and Disease. Genes (Basel) 11: 110. [PubMed: 31963661]
9. Ecsedi S, Rodríguez-Aguilera JR, and Hernandez-Vargas H. 2018. 5-Hydroxymethylcytosine (5hmC), or How to Identify Your Favorite Cell. Epigenomes 2: 3.
10. Rodríguez-Aguilera JR, Ecsedi S, Goldsmith C, Cros M-P, Domínguez-López M, Guerrero-Celis N, Pérez-Cabeza de Vaca R, Chemin I, Recillas-Targa F, Chagoya de Sánchez V, and Hernández-Vargas H. 2020. Genome-wide 5-hydroxymethylcytosine (5hmC) emerges at early stage of in vitro differentiation of a putative hepatocyte progenitor. Sci Rep 10: 7822. [PubMed: 32385352]
11. Hahn MA, Szabó PE, and Pfeifer GP. 2014. 5-Hydroxymethylcytosine: a stable or transient DNA modification? Genomics 104: 314–323. [PubMed: 25181633]
12. Ghoreschi K, Laurence A, Yang X-P, Tato CM, McGeachy MJ, Konkel JE, Ramos HL, Wei L, Davidson TS, Bouladoux N, Grainger JR, Chen Q, Kanno Y, Watford WT, Sun H-W, Eberl G,

Shevach EM, Belkaid Y, Cua DJ, Chen W, and O'Shea JJ. 2010. Generation of pathogenic T(H)17 cells in the absence of TGF-β signalling. Nature 467: 967–971. [PubMed: 20962846]

13. Wang Y, Kissenpfennig A, Mingueneau M, Richelme S, Perrin P, Chevrier S, Genton C, Lucas B, DiSanto JP, Acha-Orbea H, Malissen B, and Malissen M. 2008. Th2 lymphoproliferative disorder of LatY136F mutant mice unfolds independently of TCR-MHC engagement and is insensitive to the action of Foxp3+ regulatory T cells. J Immunol 180: 1565–1575. [PubMed: 18209052]
14. Hirota K, Duarte JH, Veldhoen M, Hornsby E, Li Y, Cua DJ, Ahlfors H, Wilhelm C, Tolaini M, Menzel U, Garefalaki A, Potocnik AJ, and Stockinger B. 2011. Fate mapping of IL-17-producing T cells in inflammatory responses. Nat Immunol 12: 255–263. [PubMed: 21278737]
15. Levéen P, Carlsén M, Makowska A, Oddsson S, Larsson J, Goumans M-J, Cilio CM, and Karlsson S. 2005. TGF-beta type II receptor-deficient thymocytes develop normally but demonstrate increased CD8+ proliferation in vivo. Blood 106: 4234–4240. [PubMed: 16131565]
16. Goldsmith C, Rodríguez-Aguilera JR, El-Rifai I, Jarretier-Yuste A, Hervieu V, Raineteau O, Saintigny P, Chagoya de Sánchez V, Dante R, Ichim G, and Hernandez-Vargas H. 2021. Low biological fluctuation of mitochondrial CpG and non-CpG methylation at the single-molecule level. Sci Rep 11: 8032. [PubMed: 33850190]
17. Leger A, and Leonardi T. 2019. pycoQC, interactive quality control for Oxford Nanopore Sequencing. Journal of Open Source Software 4: 1236.
18. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, and Timp W. 2020. Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat. Biotechnol. 38: 433–438. [PubMed: 32042167]
19. Kovaka S, Fan Y, Ni B, Timp W, Schatz MC, Fan Y, Ni B, Timp W, and Schatz MC. 2021. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. Nature Biotechnology 39: 431–441.
20. Nath AP, Braun A, Ritchie SC, Carbone FR, Mackay LK, Gebhardt T, and Inouye M. 2019. Comparative analysis reveals a role for TGF-β in shaping the residency-related transcriptional signature in tissue-resident memory CD8+ T cells. PLoS One 14: e0210495.
21. Gigante S, Gouil Q, Lucattini A, Keniry A, Beck T, Tinning M, Gordon L, Woodruff C, Speed TP, Blewitt ME, and Ritchie ME. 2019. Using long-read sequencing to detect imprinted DNA methylation. Nucleic Acids Res. 47: e46. [PubMed: 30793194]
22. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, and Ecker JR. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322. [PubMed: 19829295]
23. Korthauer K, Chakraborty S, Benjamini Y, and Irizarry RA. 2019. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. Biostatistics 20: 367–383. [PubMed: 29481604]
24. Lawson JT, Tomazou EM, Bock C, and Sheffield NC. 2018. MIRA: an R package for DNA methylation-based inference of regulatory activity. Bioinformatics 34: 2649–2650. [PubMed: 29506020]
25. Swain SL 1995. T-cell subsets. Who does the polarizing? Curr Biol 5: 849–851. [PubMed: 7583138]
26. Qiu Y, Zhu Y, Yu H, Zhou C, Kijlstra A, and Yang P. 2018. Dynamic DNA Methylation Changes of Tbx21 and Rorc during Experimental Autoimmune Uveitis in Mice. Mediators Inflamm 2018: 9129163.
27. Tsagaratou A, Äijö T, Lio C-WJ, Yue X, Huang Y, Jacobsen SE, Lähdesmäki H, and Rao A. 2014. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proc Natl Acad Sci U S A 111: E3306–3315. [PubMed: 25071199]
28. Yang B-H, Floess S, Hagemann S, Deyneko IV, Groebe L, Pezoldt J, Sparwasser T, Lochner M, and Huehn J. 2015. Development of a unique epigenetic signature during in vivo Th17 differentiation. Nucleic Acids Res 43: 1537–1548. [PubMed: 25593324]
29. Li W, and Liu M. 2011. Distribution of 5-hydroxymethylcytosine in different human tissues. J Nucleic Acids 2011: 870726.

30. Kriaucionis S, and Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324: 929–930. [PubMed: 19372393]

31. Zhu H, Wang G, and Qian J. 2016. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet 17: 551–565. [PubMed: 27479905]

32. Wang Y, Godec J, Ben-Aissa K, Cui K, Zhao K, Pucsek AB, Lee YK, Weaver CT, Yagi R, and Lazarevic V. 2014. The transcription factors T-bet and Runx are required for the ontogeny of pathogenic interferon-γ-producing T helper 17 cells. Immunity 40: 355–366. [PubMed: 24530058]

33. Zhu J, Yamane H, and Paul WE. 2010. Differentiation of effector CD4 T cell populations (*). Annu Rev Immunol 28: 445–489. [PubMed: 20192806]

34. Rasmussen A, Hildonen M, Vissing J, Duno M, Tümer Z, and Birkedal U. 2022. High Resolution Analysis of DMPK Hypermethylation and Repeat Interruptions in Myotonic Dystrophy Type 1. Genes (Basel) 13: 970. [PubMed: 35741732]

35. Nestor CE, Lentini A, Hägg Nilsson C, Gawel DR, Gustafsson M, Mattson L, Wang H, Rundquist O, Meehan RR, Klocke B, Seifert M, Hauck SM, Laumen H, Zhang H, and Benson M. 2016. 5-Hydroxymethylcytosine Remodeling Precedes Lineage Specification during Differentiation of Human CD4(+) T Cells. Cell Rep 16: 559–570. [PubMed: 27346350]

36. Martin M, Ancey P-B, Cros M-P, Durand G, Le Calvez-Kelm F, Hernandez-Vargas H, and Herceg Z. 2014. Dynamic imbalance between cancer cell subpopulations induced by transforming growth factor beta (TGF-β) is associated with a DNA methylome switch. BMC Genomics 15: 435. [PubMed: 24898317]

37. Ancey P-B, Ecsedi S, Lambert M-P, Talukdar FR, Cros M-P, Glaise D, Narvaez DM, Chauvet V, Herceg Z, Corlu A, and Hernandez-Vargas H. 2017. TET-Catalyzed 5-Hydroxymethylation Precedes HNF4A Promoter Choice during Differentiation of Bipotent Liver Progenitors. Stem Cell Reports 9: 264–278. [PubMed: 28648900]

38. Vincenzetti L, Leoni C, Chirichella M, Kwee I, and Monticelli S. 2019. The contribution of active and passive mechanisms of 5mC and 5hmC removal in human T lymphocytes is differentiation- and activation-dependent. Eur J Immunol 49: 611–625. [PubMed: 30698829]

39. Tan L, Fu L, Zheng L, Fan W, Tan H, Tao Z, and Xu Y. 2022. TET2 Regulates 5-Hydroxymethylcytosine Signature and CD4+ T-Cell Balance in Allergic Rhinitis. Allergy Asthma Immunol Res 14: 254–272.

40. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, and Meehan RR. 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res 22: 467–477. [PubMed: 22106369]

41. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, and Loose M. 2021. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. Nat Biotechnol 39: 442–450. [PubMed: 33257864]

Key Points

- Cas9-based methylation profiling of a 10-gene panel distinguishes T helper subtypes.
- Cas9 off-targets provide further indication of transcription factor activity.
- 5mCpG on the same targets distinguishes putative pathogenic Th17 cells in vivo.

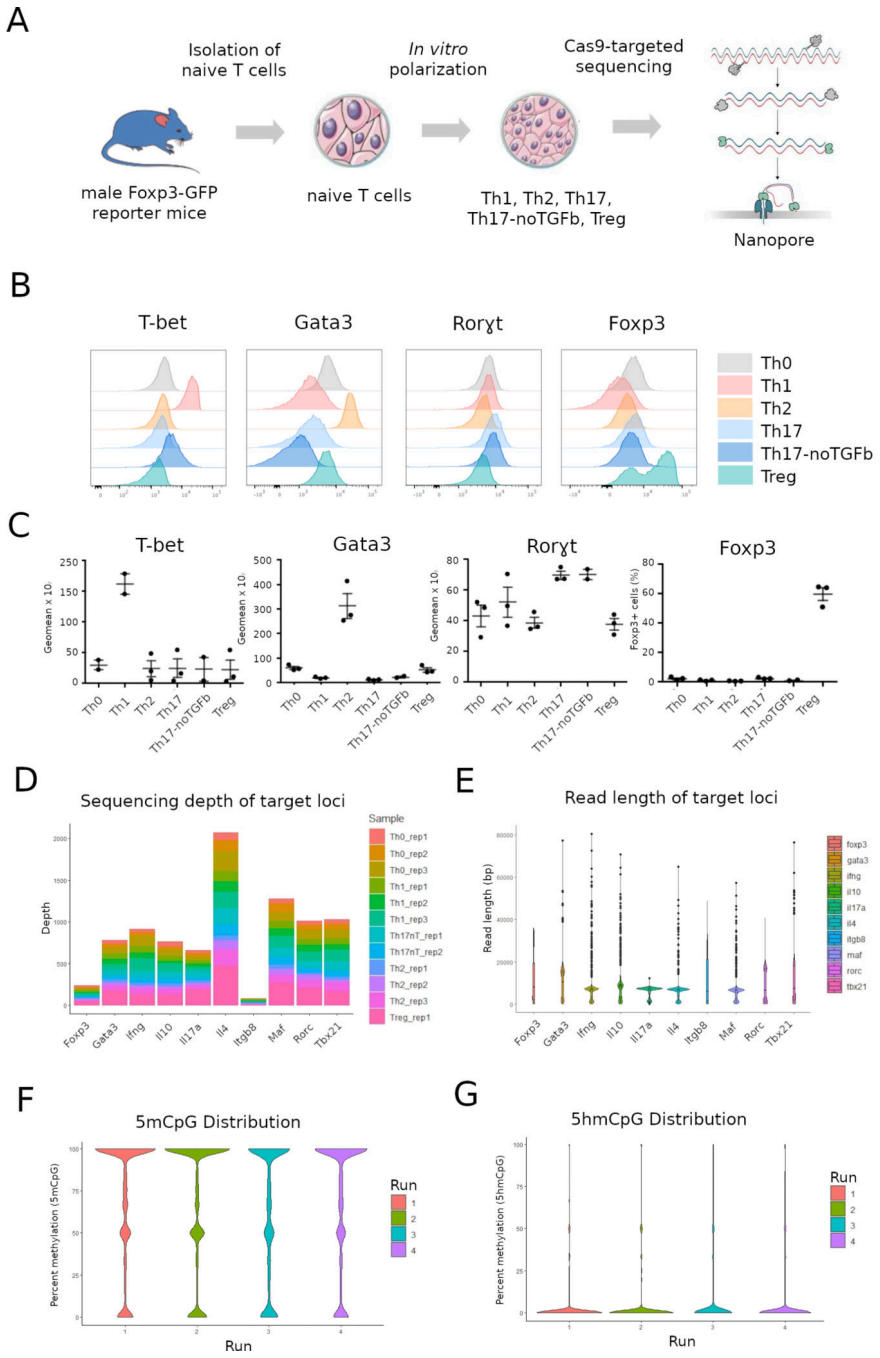

Figure 1. T-cell polarization, nanopore sequencing and validation.

A. Experimental overview. T cells were polarized *in vitro* towards Th0, Th1, Th2, Th17, Th1/17 or Treg phenotype, followed by Cas9 targeted sequencing with Nanopore. B. Transcription factor expression (ICS at day 4 of polarization) for each type of *in vitro* T cell polarization. C. Quantification of pooled independent experiments performed for transcription factor expression (ICS at day 4 of polarization). D. Nanopore sequencing coverage for each sample and each Cas9-targeted region. E. Length distribution of all reads aligning to each targeted region. F. Distribution of 5mCpG data (from 0 to 100%

methylation) across four independent sequencing runs. G. Distribution of 5hmCpG data (from 0 to 100% methylation) across four independent sequencing runs.

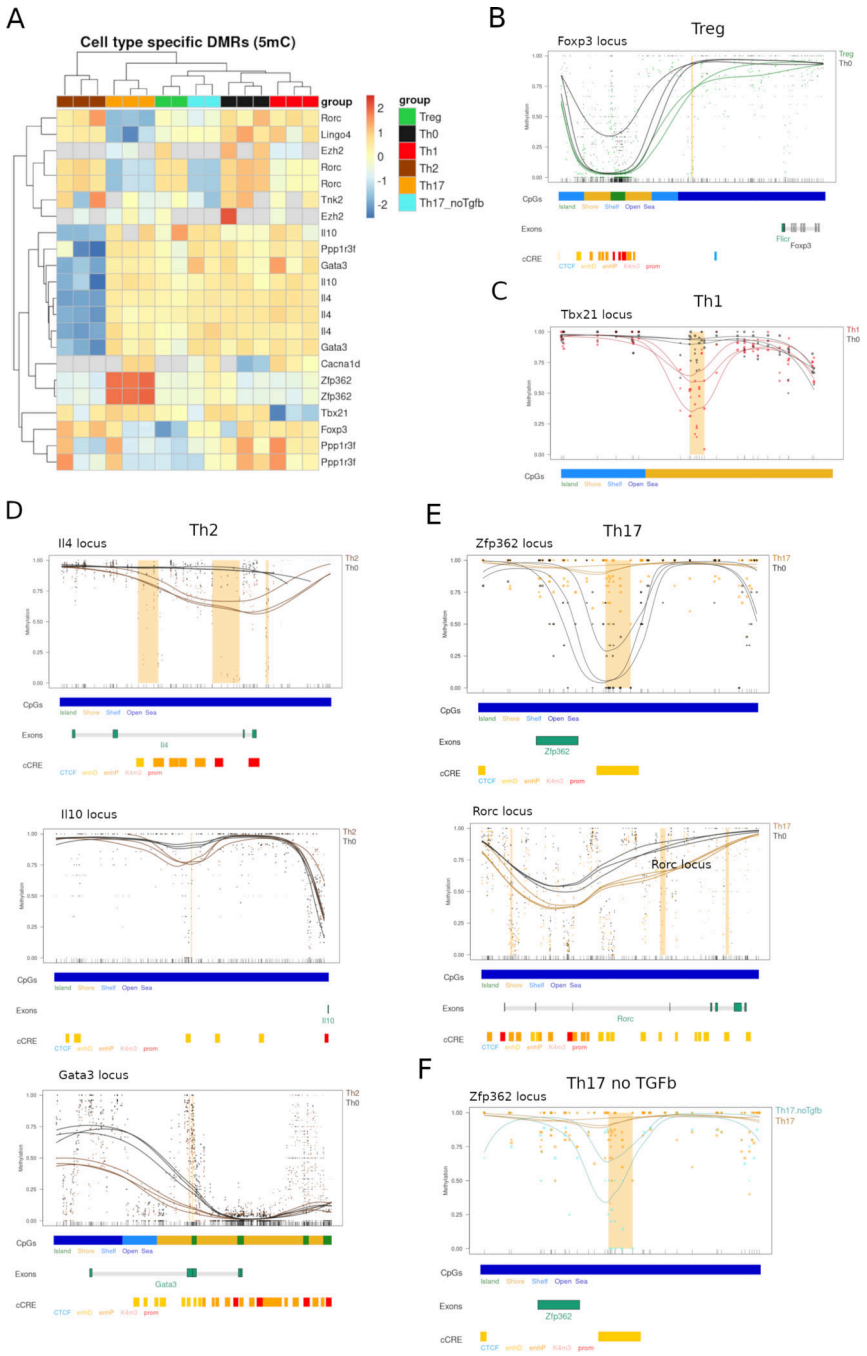

Figure 2. 5mCpG landscape of immune identity genes in CD4 Th cells.

A. Heatmap of cell type specific DMRs (5mCpG) in CD4 Th cells. For visualization, a z-score normalization was performed on the methylation values across samples for each DMR. Cell type-specific DMR plots are shown for Treg (B), Th1 (C), Th2 (D), Th17 (E) and Th17_noTGFb cells (F). Each DMR plot depicts one or several differentially methylated regions (DMR) (highlighted in orange) at the targeted locus. Methylation levels in the y axis correspond to beta values, where 0 corresponds to no methylation and 1 corresponds to 100% methylation. Two tracks below the main plot ("CpGs" and "Exons") respectively

provide information about the density of CpG sites and the overlap with annotated genes and exons. The third track corresponds to ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types. Such elements provide additional context information and were defined based on DNAse hypersensitivity and CTCF and histone binding (H3K4me3 and H3K27ac). Exact genomic positions and annotations are included in Supplemental Table 5.

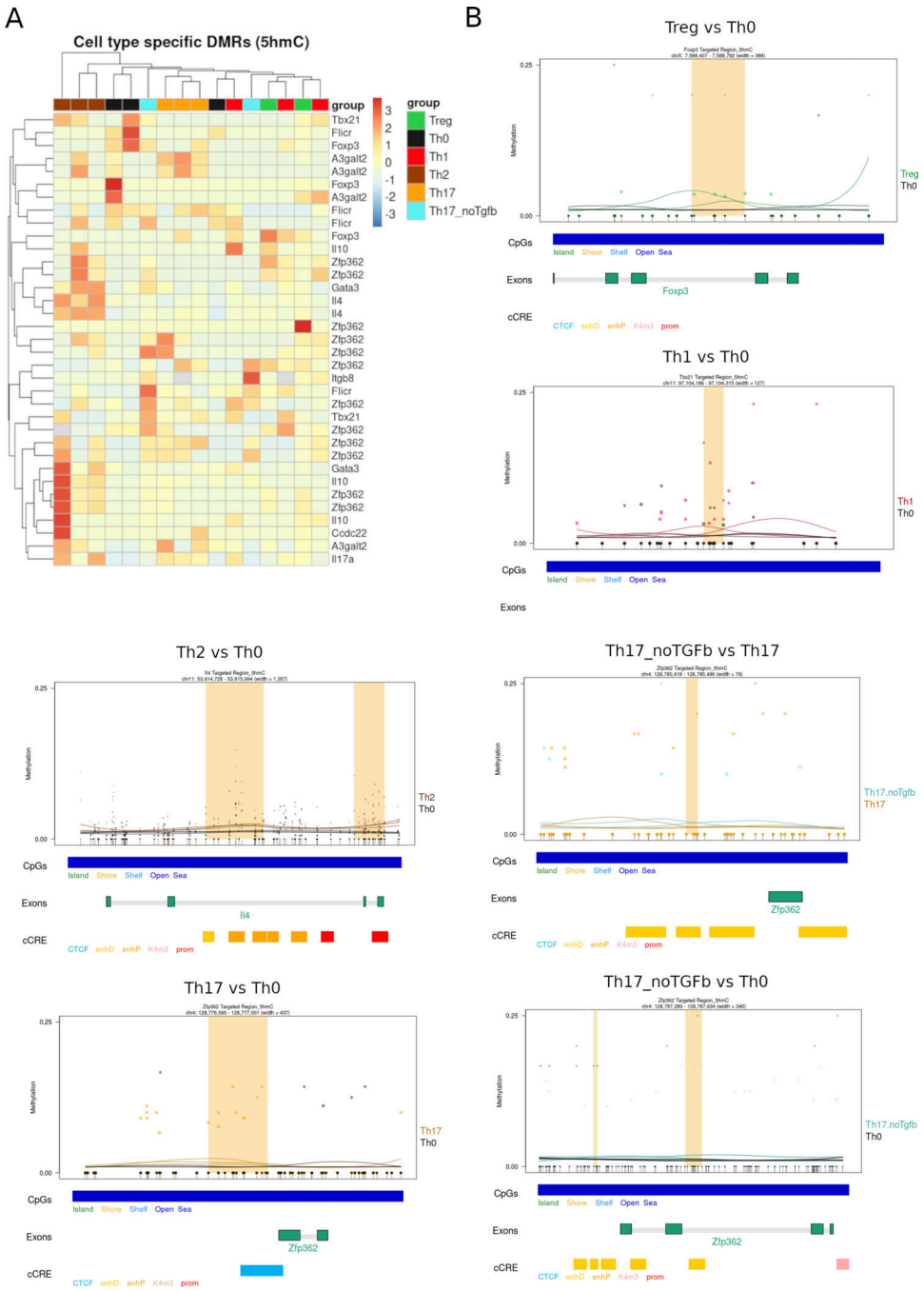

Figure 3. 5hmCpG landscape of immune identity genes in CD4 Th cells.

A. Heatmap of cell type specific regions (5hmCpG) in CD4 Th cells. For visualization, a z-score normalization was performed on the methylation values across samples for each DMR. DMRplots (B-F) and their corresponding tracks are the same as those in Fig. 2, with a shorter y axis limits (0 to 25%, instead of 0 to 100% used in Fig. 2).

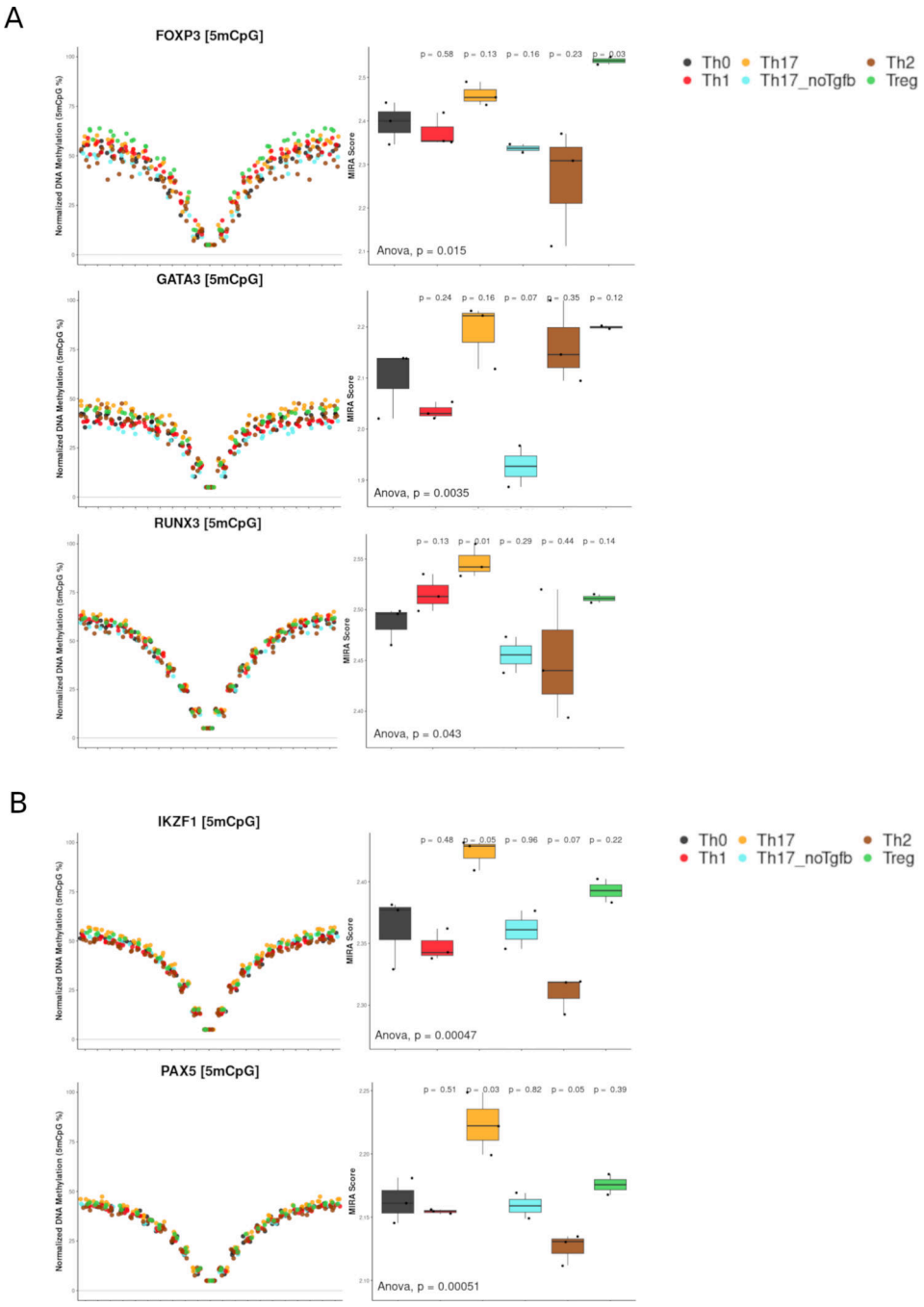

Figure 4. DNA methylation in Th cell transcription factor binding sites by MIRA.

TF activity was inferred for each sample from the binned DNA methylation distribution across annotated TF targets. For each TF, the MIRA profile is presented on the left and the aggregated score for each Th cell subset as boxplots on the right. p-values are based on t test comparisons of each subset to the reference Th0 polarized cells and ANOVA for all comparisons. (A) MIRA profiles for selected TFs. (B) MIRA profiles for unbiased TF analysis (see description on the text).

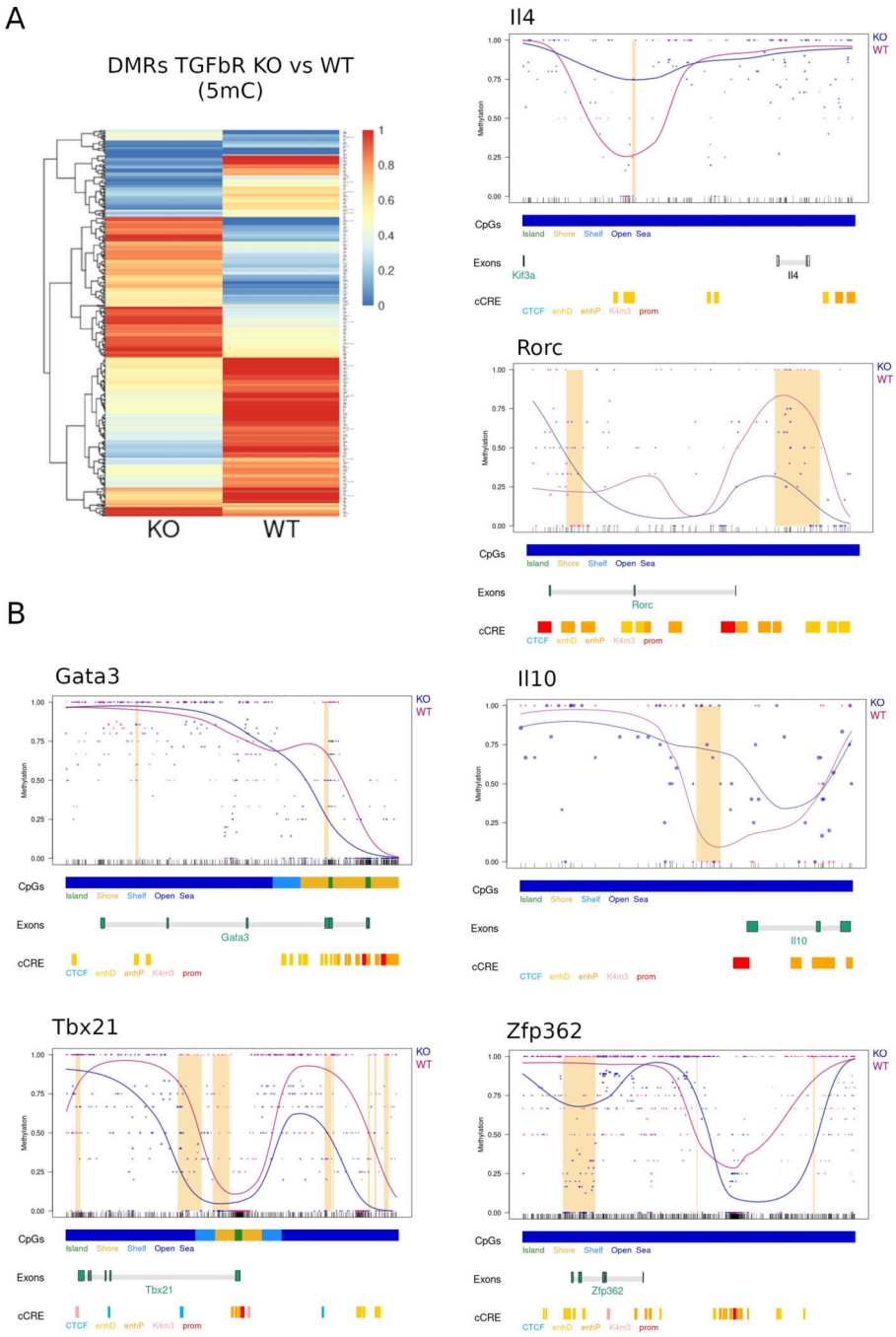

Figure 5. In vivo detection of 5mCpG in a mouse model of pathogenic Th17 cells.

A. Heatmap of average 5mCpG methylation for all differentially methylated regions (DMRs) detected in TGFbR-KO cells relative to WT cells. B. DMR plots for selected regions distinguishing TGFbR-KO cells form their WT counterpart. Tracks and plot were as described in Fig. 2. Exact genomic positions and annotations are included in Supplemental Table 7.