SUPPLEMENTARY INFORMATION

Supplementary methods

Peak saturation analysis

Peak saturation curves for MAP-seq and ChIP-seq experiments were carried out in a similar manner to that described in (Kharchenko et al. 2008). Briefly, for each of Th1 MAP-seq, Th1 H3K4me3 ChIP-seq and Th1 H3K27me3 ChIP-seq samples, random proportions of reads were selected to form subsets of aligned reads (5%, 10%, 15%...95%). Background subtraction (Th1_MAP only) and normalisation (using the scaling factor from the original analysis) were carried out followed by peak finding using the original parameters (see Table S1). Peaks from these subsets of data were intersected with the complete set of peaks to determine what fraction of peaks was being detected. As this work predominantly focuses on CGIs, peaks from subsets of data were intersected with the complete set of mouse CGIs to determine if the fraction of CGI peaks detected was approaching saturation.

Supplementary figures

Figure S1. MAP-seq interrogates CpG-rich regions.

(A) The CpG density of regions identified by MAP-seq in a typical experiment (CD4 cell MAP-seq) calculated from peak-called data. The 10^{th} percentile of the data (3.3 CpGs per 100bp) is taken to represent the CpG density necessary for detection of a region by MAP-seq.

Figure S2. Verification of MAP-seq results by bisulfite sequencing.

Examples of differentially methylated CGIs that were verified by bisulfite sequencing. For each example: Upper panel; MAP-seq read density profiles (red), CGIs (blue), bisulfite amplicon (grey), differentially methylated CGIs are asterisked and the arrow indicates the origin and direction of transcription. Middle panel; bisulfite sequencing results for the amplicon shown above. Filled circles represent methylated CpGs and empty circles represent unmethylated CpGs. Lower panel; DNA methylation frequency of each clone analysed by bisulfite to look for evidence of composite or allelic methylation. Equal numbers of clones from each cell type were chosen randomly for analysis. CGIs at the following loci are shown and genomic location is indicated: *(A) Npb (B) Tnfaip2 (C) Kcnn4 (D) Sema4a (E) Dtx1 (F) Nr4a3 (G) Hic1 (H) Klf2* and *(I) Gata3*.

Figure S3. Location of cell type-specific methylation with respect to alternative CGI sets.

The fraction of the genome that can be interrogated by MAP-seq was categorised as overlapping a CGI (CGI), not overlapping but within 2kb of a CGI (0-2kb) or greater than 2kb from a CGI (>2kb) using either the mouse UCSC CGI set *(A and B)* or a model-based definition of CGIs (Wu et al. 2010) *(C and D).* The location of cell or tissue-specific DNA methylation changes was then expressed as the number of methylation changes occurring per kilobase (kb) of genome in each category. *(A)* All DNA methylation changes occurring between different immune cell types (UCSC CGIs) *(B)* DNA methylation changes occurring between brain and CD4 cells (UCSC CGIs). *(C)* All DNA methylation changes occurring between different immune cell types (model-based CGIs) *(D)* DNA methylation changes occurring between brain and CD4 cells (model-based CGIs).

Figure S4. At differentially methylated CGIs the greatest change in methylation is over the CGI and not the 2kb flanks.

The number of sequencing hits over the CGI and its 2kb flanks was calculated for all CGIs showing differential methylation between brain and CD4 cells (Brain vs CD4) and dendritic cells and Th1 cells (DC vs Th1). This was done using a sliding window analysis to count the number of sequencing hits per base in a 100bp window (20bp slide; flanks) or 12% window (2% slide; CGI). CGIs were divided into those showing increased and decreased methylation. The average number of hits per base across these CGIs and their flanks was plotted by taking the median values obtained from the sliding window analysis.

(A) CGIs showing increased methylation in brain vs. CD4 cells. *(B)* CGIs showing decreased methylation in brain vs. CD4 cells. Brain (blue), CD4 cells (red). *(C)* CGIs showing increased methylation in DC vs. Th1. *(D)* CGIs showing decreased methylation in DC vs. Th1. Dendritic cells (green), Th1 cells (purple).

Figure S5. Intragenic CGIs show preferential tissue-specific methylation between brain and CD4 cells.

The percentage of each class of CGI (TSS, Intragenic, Intergenic) showing a DNA methylation change between brain and CD4 cells.

Figure S6. H3K27me3 ChIP-seq

H3K27me3 ChIP-seq confirms enrichment for H3K27me3 at the positive control gene *(A) Sox2* but not at the negative control gene *(B) Actb* which is heavily enriched for H3K4me3. Both *(C) Tnfaip2* and *(D) 2810459M11Rik* show cell type-specific CGI methylation, lack H3K4me3 but are enriched for H3K27me3. Read density profiles are shown for MAP-seq (red), H3K27me3 ChIP-seq (cyan) and H3K4me3 (green). CGIs are indicated (blue) and differentially methylated CGIs are asterisked.

Figure S7. A small number of differentially methylated intragenic CGIs are bivalent. *(A)* Differentially methylated intragenic CGIs in Th1 and dendritic cells that are H3K4me3+ and H3K27me3+ in both cell types (4 out of 63 CGIs). Bivalent intragenic CGIs in *(B) Phospho1* and *(C) Kcnk12*. Read density profiles are shown for MAP-seq (red), H3K27me3 ChIP-seq (cyan) and H3K4me3 (green). CGIs are indicated (blue) and differentially methylated CGIs are asterisked.

Figure S8. H3K4me3 at intragenic CGIs in ES cells and dendritic cells.
(A) The percentage of TSS and intragenic CGIs that are positive for H3K4me3 in ES cells (grey), dendritic cells (black) or both (stripes).

Figure S9. Peak saturation analysis for high throughput sequencing experiments. Random fractions of sequencing reads were selected (5%, 10%, 15%...95%) and plotted against the fraction of peaks (reference calls) detected at that sequencing depth (blue line). Fractions of sequencing reads were also plotted against the fraction of CGI peaks detected (red line) as CGIs are the primary focus of this work. Peak saturation curves are shown for *(A)* Th1 MAP-seq, *(B)* Th1 H3K4me3 ChIP-seq and *(C)* Th1 H3K27me3 ChIP-seq experiments. For CGI-associated peaks, sequencing is approaching saturation.

Supplementary tables

Table S1. Sample list and analysis parameters.
See attached file “Deaton_Table S1.xls”.

Table S2. CGIs showing differential methylation in the immune system.
CGIs identified as differentially methylated are given for all immune cell comparisons carried out. The co-ordinates of the CGI (as identified by Illingworth et al, 2010) and the co-ordinates of the differentially methylated region (DMR) are given as well as

information about whether the region gains (+1) or looses (-1) methylation in the first cell type compared to the second and any overlapping RefSeq genes.
See attached file “Deaton_Table S2.xls”.

Table S3. Bisulfite verification of cell type-specific methylation.
Differentially methylated CGIs identified by MAP-seq that were verified by bisufite sequencing. The gene in which each CGI is located along with the genomic coordinates of the differentially methylated region is indicated. For MAP-seq, the cell type showing a decrease in methylation is indicated in brackets and for bisulfite the percentage methylation and the overall fold change in methylation is given.

Gene	**Differentially methylated region**	**MAP-seq (hypomethylated cell type in brackets)**	**Bisulfite (% methylation)**	**Fold change bisulfite**
Npb	chr11: 120469741 - 120470220	Th1 (vs DC)	54% Th1, 9% DC	6
Tnfaip2	chr12: 112683701 - 112684060	DC (vs Th1)	66% DC, 26% Th1	2.53
Kcnn4	chr7: 25161961 - 25162900	DC, B cell, CD4 (vs Th1)	66% DC, 70% B, 32% CD4, 20% Th1	3.3
Sema4a	chr3: 88240621 - 88241160	CD4, Th1, Th2 (vs DC)	80% Th1, 29% DC	2.75
Dtx1	chr5: 121160061 - 121160460	Th1 (vs CD4/DC/B)	55% Th1, 30% CD4, 17% DC, 13% B	3.23
Nr4a3	chr4: 48064301 - 48065140	Th1 (vs DC)	73% Th1, 16% DC	4.56
Hic1	chr11:74979221 - 74979660	Th1 (vs DC)	33% Th1, 7% DC	4.71
Klf2	chr8: 74844581 - 74845040	DC (vs Th1)	80% DC, 47% Th1	1.7
Gata3	chr2: 9795943 - 9796614	Th1 (vs Th2)	72% Th1, 54% Th2	1.33

Table S4. Ontology analysis of differentially methylated CGIs
Gene ontology terms enriched in genes showing differential CGI methylation in the immune system (top) and between brain and CD4 cells (bottom). All CGI-associated genes were used as a reference set and a p-value ≤ 0.05 was taken to indicate significant enrichment (for more details see “methods”). See attached file “Deaton_Table S4.xls”.

References

Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature biotechnology* **26**(12): 1351-1359.

Wu H, Caffo B, Jaffee HA, Irizarry RA, Feinberg AP. 2010. Redefining CpG islands using hidden Markov models. *Biostatistics (Oxford, England)* **11**(3): 499-514.