Supplemental Figure Legends

Figure S1. 5hmC dot blots on oligonucleotides with varying amounts of 5hmC.

Figure S2. Characterization of defined 5hmC peaks.
(a) Distribution of sizes of 5hmC regions. Mean=1571.0bp; median=1249bp; standard deviation=1200.1bp; maximum=25749bp; minimum=249bp.
(b) Distribution of distances between 5hmC regions. Mean=192.56kb; median=45kb; standard deviation=760.48kb; maximum=31468.5kb; minimum=1kb.
(c) Distribution of 5hmC regions in 24 chromosomes.

Figure S3. Correlation of 5hmC and gene expression.
(a) 5hmC over genes. 5hmC peak density was plotted over RefSeq genes in 500 base-pair windows.
(b) Number of RefSeq annotations with given log10(RPKM) values were plotted. “Genes with 5hmC region in body” were genes that contained at least one 5hmC region between the annotated transcription start and termination sites. “Genes with 5hmC region in promoter” were genes that contained at least one 5hmC region between the annotated transcription start and 5kb upstream (i.e. 5’ end) of the transcription start site.

Figure S4. 5hmC and different classes of putative enhancers.
(a) Percentage of active (Class I) and poised (Class II) enhancers[12] overlapping with 5hmC regions. Random regions with the same number and size distribution as the putative enhancers were generated and overlap with 5hmC regions was calculated 100 times. Error bars represent standard deviation.
(b) Percentage of 5hmC marked enhancers overlapping with Class I and Class II enhancers. Random regions with the same number and size distribution as the 5hmC marked enhancers were generated and overlap with putative enhancers was calculated 100 times. Error bars represent standard deviation.

**Figure S5. GO analysis of genes that overlap with 5hmC peaks.**

Gene Ontology biological processes analysis of genes that are associated with 5hmC regions using GREAT[13]. The y-axis corresponds to the $-\text{log}10$ of the binomial raw P-values.

**Figure S6. Genome-browser views of 5hmC, enhancers and TFBS.**

(a) 5hmC over genes specifically expressed in human ES cells[15]. hmeDIP-seq read densities (reads/bp/million uniquely mapping reads) were plotted in 200bp windows. Predicted enhancers[9] and TFBS[14] are also shown.

(b) 5hmC over control genes.

**Figure S7. Sequence composition over 5hmC regions in different genomic locations.**

(a) Base composition of the Watson strand over the 5’ and 3’ boundaries of 5hmC regions overlapping with RefSeq genes.

(b) Base composition over 5hmC regions overlapping with predicted hESC enhancers.

(c) Base composition over 5hmC regions neither overlapping with RefSeq genes nor predicted enhancers.
Figure S1

- C: 6 CpGs
- mC: 3 CpGs
- hmC: 1 CpG

8 pmoles hmC
Figure S2

(a) Size distribution of defined 5hmC peaks

(b) Inter-peak distance distribution between defined 5hmC peaks

(c) hmeDIP peak distribution

- hmeDIP peaks
- genome size

chromosome
Figure S3

(a) 5hmC peak density distribution around TSS and TTS.

(b) Histograms showing the distribution of log10(RPKM) for different gene categories.

- All genes
- Genes with 5hmC peak in body
- Genes with 5hmC peak in promoter
Figure S4

(a) % overlap of putative enhancers with 5hmC peaks

(b) % overlap of 5hmC marked enhancers with putative enhancers
Figure S5

Gene Ontology Biological Process

-log10(P)

embryonic pattern specification
embryonic cortex morphogenesis
cerebellar cortex morphogenesis
positive regulation of developmental growth
non-canonical Wnt receptor signaling pathway
negative regulation of transforming growth factor beta receptor signaling pathway
mesonephros development
phasic smooth muscle contraction
Wnt receptor signaling pathway, calcium modulating pathway
paraxial mesoderm development
SMAD protein signal transduction
Figure S6

(a) Active Motif hmeDIP-seq

(b) Active Motif hmeDIP-seq

Input DNA
NANOG
OCT4
hESC enhancer

Diagenode hmeDIP-seq

Input DNA
NANOG
OCT4
hESC enhancer

0.005
0
0.005
0
0.005
0
0.005
0

GDF3
ELA1
LOC196415

50,010,000 50,020,000
25,035,000 25,040,000
Figure S7

(a) Fraction of bases vs. Distance from 5hmC region boundary (bp)

(b) GC skew vs. Distance from 5hmC region boundary (bp)

(c) Fraction of bases vs. Distance from 5hmC region boundary (bp)

Distance from 5hmC region boundary (bp)
| 5hmC enriched regions | Chr1: 11321049 | CTCTTCCTCCTCCTCGTCCT |
| | Chr1: 149533282 | CATCCAGGAACCCACAGTCCT |
| | Chr1: 149536730 | CATCACCTGGAAAGGAAGA |
| | Chr1: 83930958 | CCCTGCAAATTAGGGATGAA |
| | Chr1: 232733772 | ACAGAGAACGCCACTGACT |
| | Chr1: 232733772 | TTGCAGGAGTGAAGTGTG |
| control regions | Chr1: 149536730 | GCTCCTGGTCTTGGCTTTGAC |
| Chr2: 135171597 | CCAAGCCCAGCTAATGTATG |
| | Chr2: 135171597 | CGCCTAATCCCAAGCTATTTG |
| | Chr2: 135171597 | CCAGCCTGGTTGGCTAAAAGT |
| | Chr2: 135171597 | TGGTGGCAGGCACCTGTA |