SUPPLEMENTAL INFORMATION

GO term analysis of differentially methylated SUMIs. GO term analysis of the 458 SUMIs with the largest differential methylation between human and chimp shows that they are more frequently associated with proteins localized to the plasma membrane, compared to all orthologous SUMIs (Table S3).

Supplemental Figures and Figure Legends

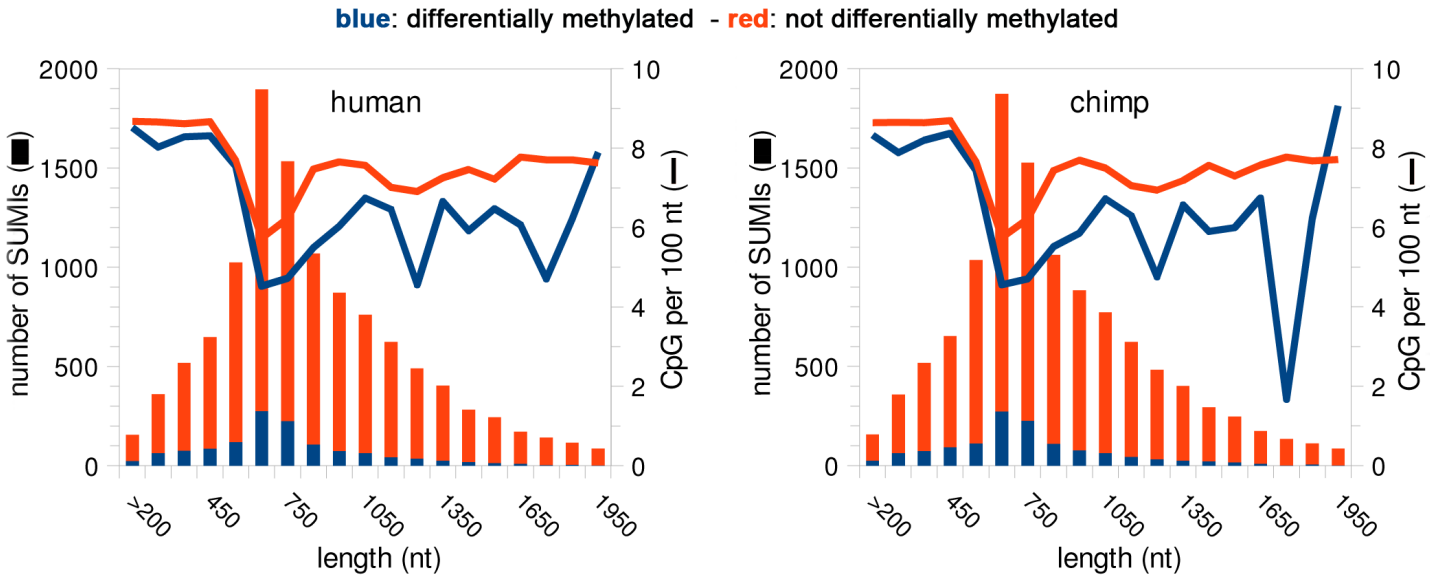

Figure S1. Length distribution and average CpG content of orthologous human-chimp SUMIs and differentially methylated SUMIs. The X axes show the lengths human (left) and chimp (right) SUMIs; the number of SUMIs of a given length is shown on the left-hand Y axes. The length distribution of SUMIs is similar between the two species, and between SUMIs that are differentially methylated (blue bars) and those that are not (red bars). This length distribution is also similar to that of computationally defined CG islands, as expected because these sets largely overlap (Singer et al. 2010). CG content is reported as number of CG per hundred nucleotides (scale on right-hand Y axes). The CG content of differentially methylated SUMIs (continuous blue line) is slightly lower than in SUMIS that are not differentially methylated (red lines).

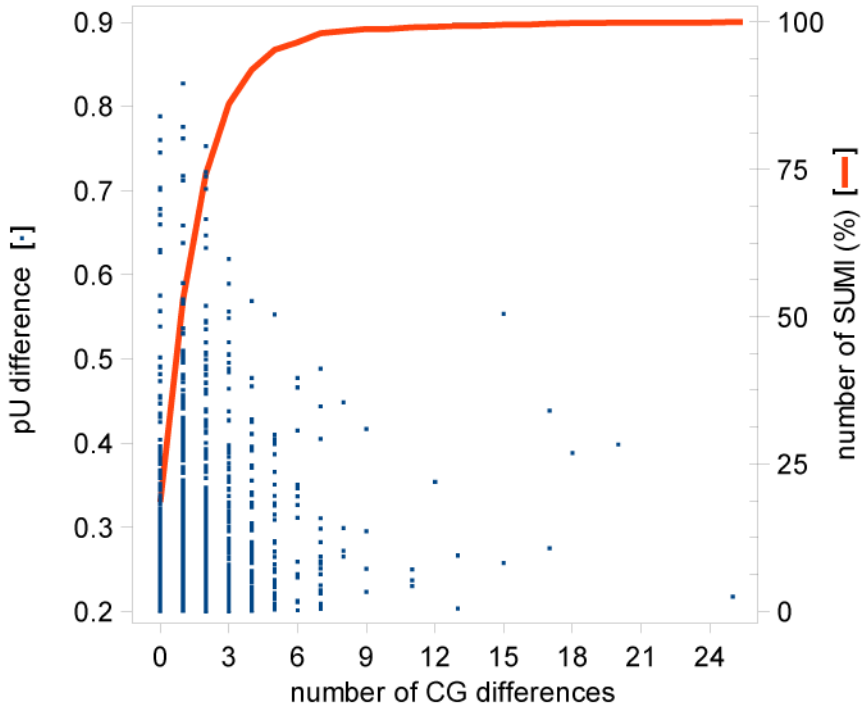

Figure S2. Differential methylation in human and chimp SUMIs is not related to gain or loss of CpGs in one of the two species. The blue dots indicate individual differentially methylated SUMIs (using the threshold of 0.2 to determine differential methylation), with the pU difference shown on the left-side Y axis and the number of CG differences on the X axis. The red line (scale on right-side Y axis) represents the running sum of the number of differentially methylated SUMIs with a number of CG differences between human and chimp smaller or equal to the value indicated on the X axis. ~20% of the differentially methylated SUMIs have 0 CG differences, and ~50% have 0 or 1 CG differences. SUMIs with greater differential methylation do not have a larger number of CpG differences. Most SUMIs with the largest differential methylation have a small number of CG differences.

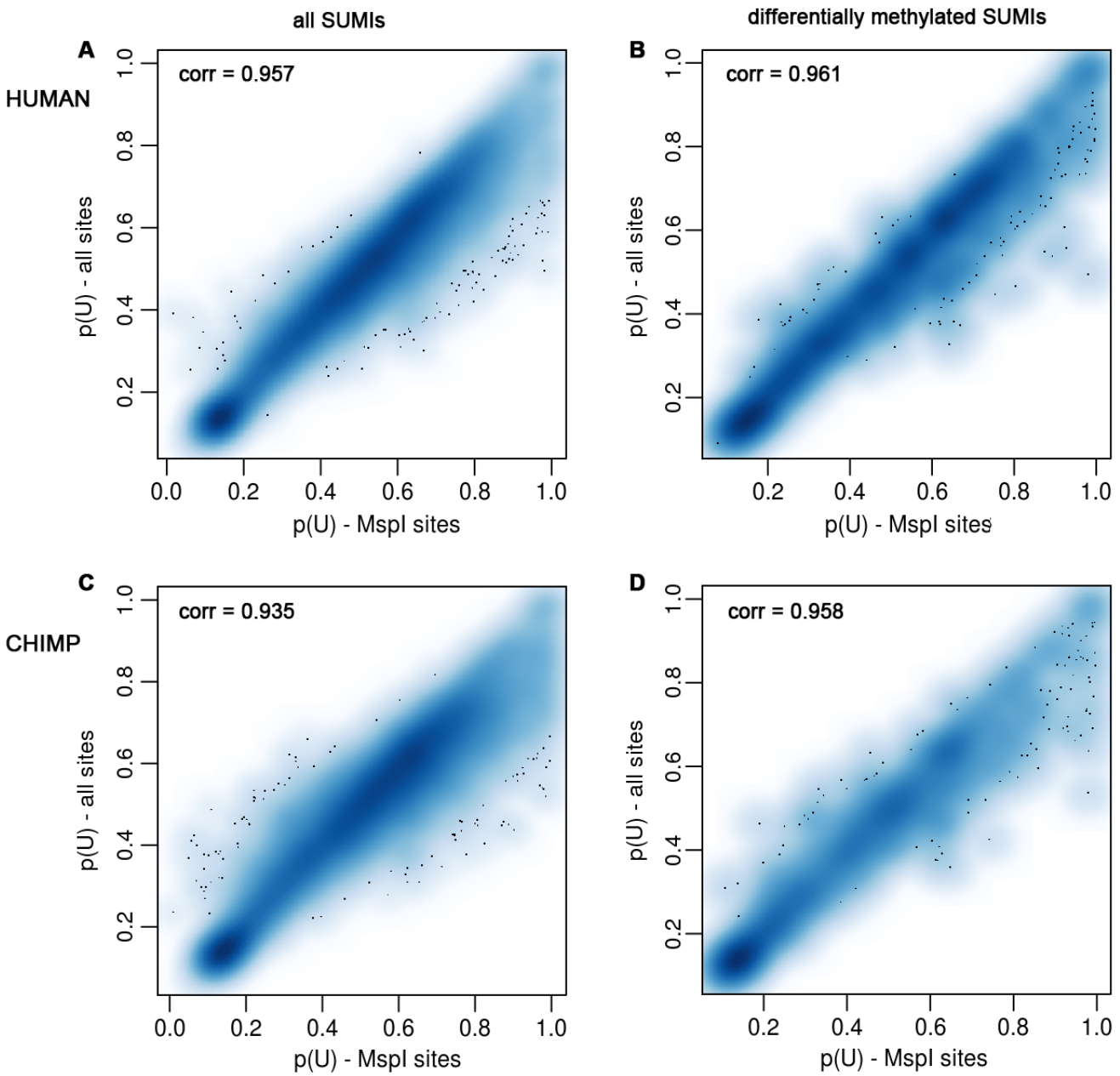

Figure S3. Comparison of data obtained by analyzing all HpaII sites with data obtained only from HpaII sites that were confirmed by MspI digestion. Digestion with HpaII's methylation-insensitive isoschizomer MspI demonstrates that a HpaII site is present in the individual analyzed, and that failure to digest with MspI is not due to absence of the restriction site. In both human (top) and chimp (bottom), there is strong correlation between the probability of methylation computed by MetMap for SUMIs obtained from HpaII sites that were confirmed by MspI digestion (X axes) and SUMIs obtained from all HpaII sites within the scope of the experiment (Y axes). Correlation is slightly stronger for differentially methylated SUMIs (B and D) than for all SUMIs (A and C).

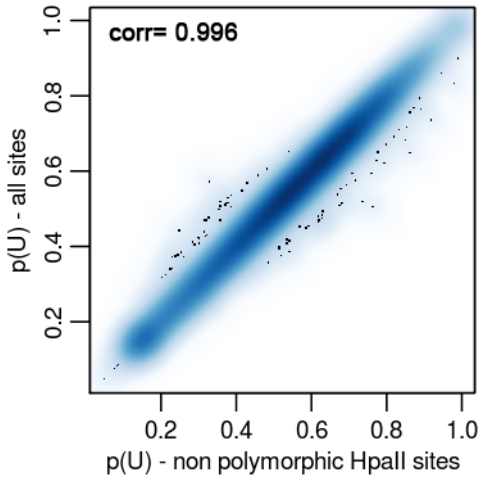

Figure S4. Human sequence polymorphism at HpaII sites does not affect the calculated methylation state of SUMIs. Comparison of methylation states of SUMIs calculated only from HpaII sites (X axis) that did not overlap with human sequence polymorphism with methylation states calculated using all HpaII sites. The strong correlation suggest that our analysis of SUMI methylation is unlikely to be affected by HpaII site polymorphism. Human sequence polymorphism data were obtained from dbSNP v131.

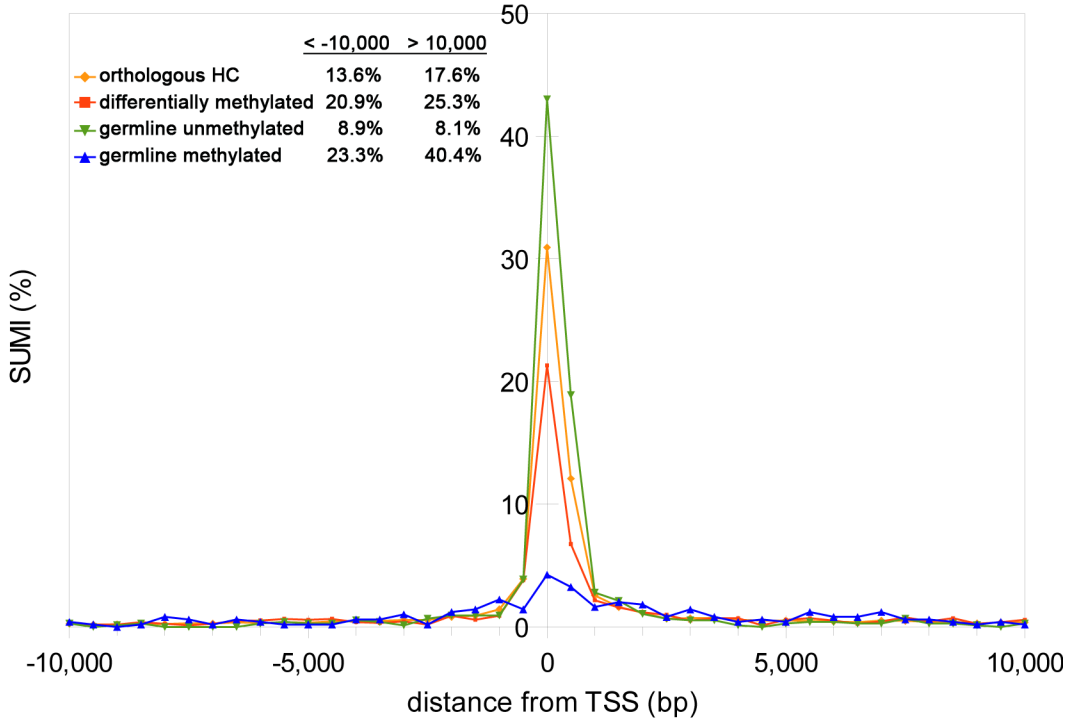

Figure S5. Distance of SUMIs from the nearest transcription start site (TSS). The plot is calculated for all orthologous human-chimp SUMIs (orange), SUMIs that are differentially methylated in human and chimp(red), SUMIs predicted to be unmethylated in the germline (green), and SUMIs predicted to be germline methylated (blue). In general, SUMIs are most likely to be close to a TSS, as expected since SUMIs usually overlap with computationally defined CG islands, and many CG islands are promoters. The proportion of SUMIs not displayed in the window containing the X axis is shown in the table insert. Differentially methylated SUMIs are slightly less likely to be in proximity to a TSS than orthologous SUMIs in general. Germline unmethylated SUMIs are usually proximal to a TSS. In contrast, germline methylated SUMIs show only a minor enrichment near TSS, and most are >10kb distant. In conjunction with the data in Table 1, this may indicate that differentially methylated and germline methylated SUMIs (CG islands) are more likely to be enhancers. Human gene annotation for hg18 was from the refGene table of the UCSC Genome Browser.

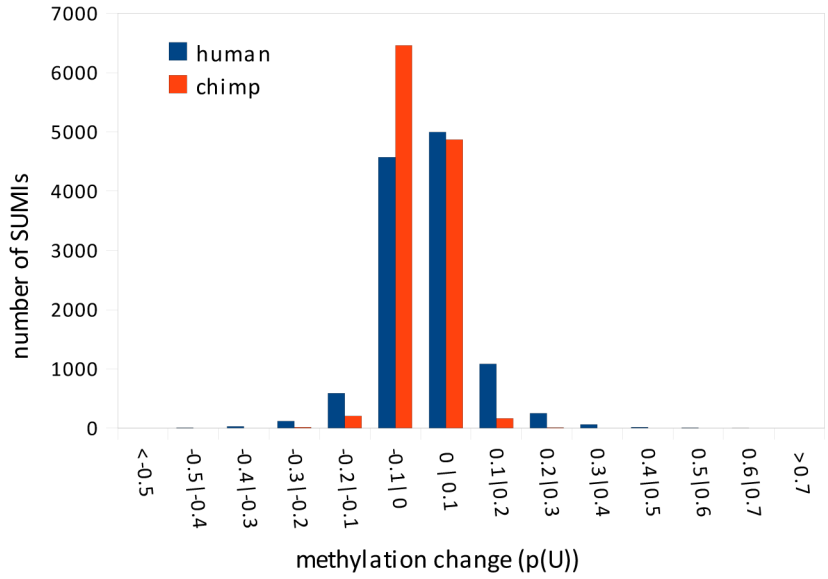

Figure S6. Distribution of methylation change, relative to the last common ancestor, in human and chimp SUMIs. The amount of change in methylation state of a SUMI was calculated as the difference between the p(U) values of the common ancestor, estimated from human, chimp and orang methylation states, and the extant species. The value is positive if the SUMI in the extant species is less methylated than the common ancestor, and negative if it is more methylated. The distributions of the amount of methylation change are shown for human (blue) and chimp (red). A larger number of human SUMIs than chimp SUMIs shows more extreme changes in methylation (SUMIs at the tails of the distribution).

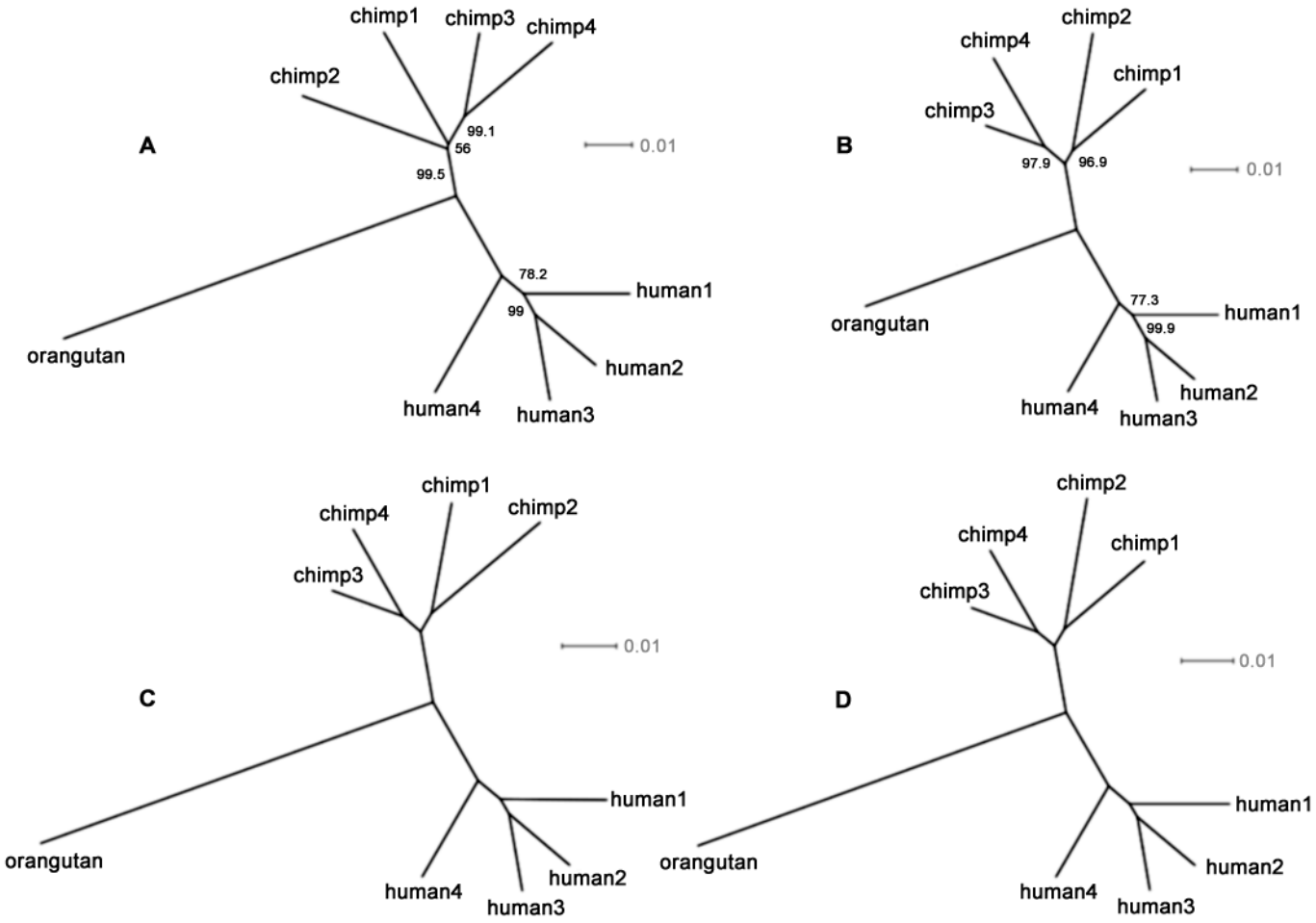

Figure S7. Phylogenetic trees built from different measures of methylation state recapitulate ape phylogeny. The tree shown in Figure 2 was constructed from the methylation states of all orthologous (human-chimp-orang) SUMIs. The trees in this figure were built using Neighbor-Joining as described in the main text. Bootstrapping values are from 1,000 replications via the "SplitsTree" program; values are shown in the figure only if they are lower than 100, i.e. all unlabelled branches have a bootstrap value of 100. The scale bars for each tree indicate the number of substitutions per site. The tree in **A** was built from the 6021 SUMIs in which the human and chimp have the same number of CCGG sites, and demonstrates that the tree built from orthologous SUMIs is not an artifact due to changes in the number of HpaII sites in the genomes of human and chimp. The tree in **B** was built from only those HpaII sites that have been confirmed by MspI digestion and deep sequencing, demonstrating that use of validated sites only obtains the same phylogeny. Tree **C** is built from the set of HpaII sites that do not have sequence polymorphisms reported in dbSNP131, producing the same conclusion as B. The tree in **D** is built from all orthologous HpaII sites in human, chimp, and orang, irrespective of their location within a SUMI or not. Consistent with the SUMI-based tree shown in Figure 2, this tree recapitulates the established phylogeny of the three species.

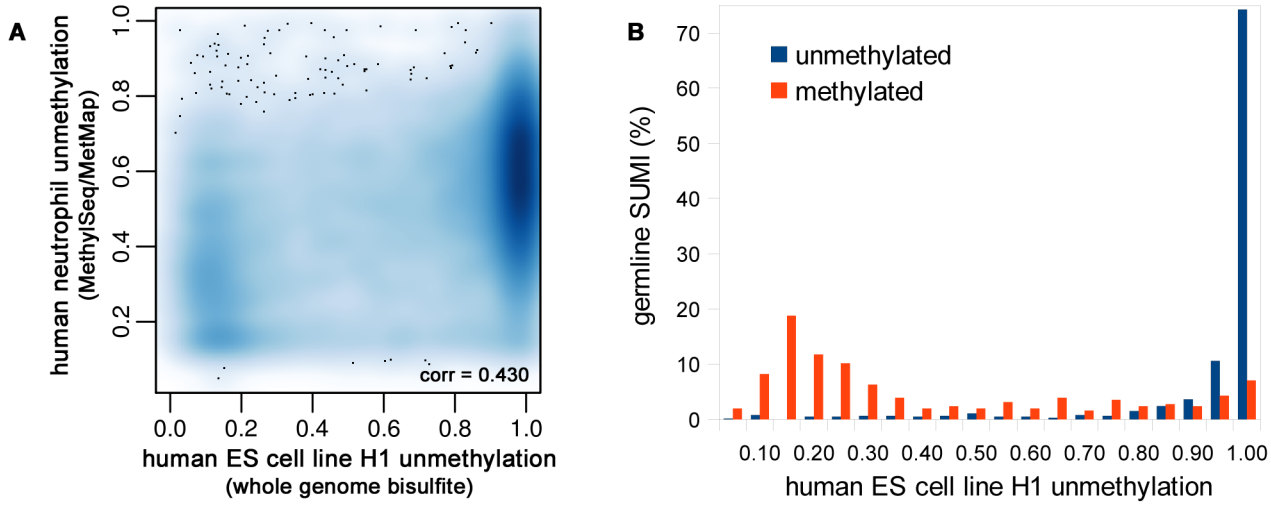

Figure S8. Correlation between SUMI methylation states in the neutrophil and the human embryonic stem cell H1. **A.** Methylation states of human SUMIs in the human embryonic stem cell H1 were calculated from the single-nucleotide resolution dataset obtained by bisulfite sequencing by Lister et al. (Lister et al. 2008) and compared to the corresponding methylation states in the neutrophil obtained with MethylSeq-MetMap. The methylation values for H1 embryonic stem cells are plotted on the X axis, and the methylation values for the neutrophil are on the Y axis. The SUMIs calculated from bisulfite sequence data in ES cell are largely fully unmethylated (right side) while a smaller proportion are fully methylated (left side) methylated. A proportion of SUMIs are methylated in both ES cell and neutrophil (bottom left). Many SUMIs that are either methylated or unmethylated in ES cells have intermediate values in the neutrophil; this may reflect either true differences between the two cell types, or the limited ability of whole-genome bisulfite sequencing to detect intermediate states because of the depth of sequencing required to detect such states. **B.** Comparison of methylation states in the human embryonic stem cell H1 and germline methylation inferred by CG decay analysis. We calculated the methylation state in the human embryonic stem cell H1 of all SUMIs that are consistently methylated (red) or unmethylated (blue) in neutrophils of human, chimp and orang. The calculated methylation state is shown on the X axis (0 = methylated, 1 = unmethylated), and the percent of SUMIs with a given methylation state is on the Y axis. Almost all SUMIs inferred to be unmethylated in the germline by CG decay are unmethylated in the H1 ES cell line. Of the SUMIs predicted to methylated in the germline, 61% are methylated in H1 cells (unmethylation value ≤ 0.35); this value is very close to the predicted fraction of methylated SUMIs of ~0.66 discussed in relation to Figure 4.

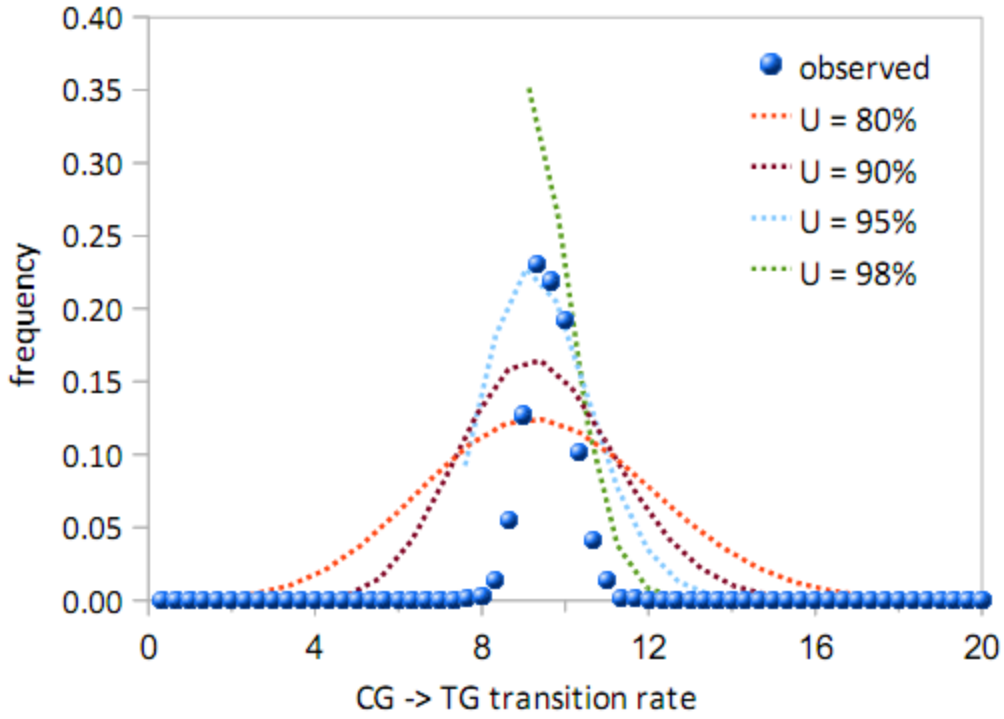

Figure S9. Estimation of the proportion of SUMIs that are unmethylated in the germline from CG decay analysis. To estimate the proportion of germline-unmethylated SUMIs in the 775 SUMIs that are consistently unmethylated in human, chimp and orang neutrophils, we calculated CG decay rates in 1000 random subsets of these SUMIs. The observed distribution of decay rates of the subsets is shown by the blue circles. Simulated distributions calculated for different proportions of unmethylated SUMIs are shown by the dotted lines. The variance of the distribution is a measure of the proportion of methylated and unmethylated SUMIs in the 775 SUMIs: a small variance indicates that most of the SUMIs are unmethylated (curve for unmethylated=98%); a large variance indicates that some of the SUMIs are methylated (curve for unmethylated=80%). The data indicate that almost all of the 775 SUMIs are unmethylated in the germline.

	Human1	Human2	Human3	Human4	Chimp1	Chimp2	Chimp3	Chimp4	Orang
Reads passing QC	6,227,749	6,202,253	6,233,646	5,067,711	6,438,740	3,908,393	4,837,176	3,402,589	11,444,363
Unique alignments	4,160,629	4,221,245	4,225,504	4,066,062	3,048,845	1,750,769	3,481,281	2,414,202	5,204,114
Unique at HpaII sites	3,976,345	4,079,770	4,072,977	3,958,110	2,741,144	1,586,822	3,226,155	2,241,081	4801233

Table S1. Summary of sequencing data used in this study. The table shows the number of reads passing the quality filter (Illumina Pipeline Chastity filter set at 0.6) that were collected and further processed for each sample ("Reads passing QC"); the number of reads with unique alignments against their respective genomes ("Unique alignments"); and the number of unique alignments at HpaII sites ("Unique at HpaII sites"). Only unique alignments at HpaII sites were processed by MetMap. The sequence data have been deposited in the NCBI GEO database, with accession number GSE22376.

	Human1	**Human2**	**Human3**	**Human4**	**Human**
SUMIs	16,904	17,596	18,179	18,700	20,986
not CGI	2,832	2,587	2,905	3,425	4,651

	Chimp1	**Chimp2**	**Chimp3**	**Chimp4**	**Chimp**
SUMIs	19,953	17,799	17,973	17,331	21,370
not CGI	4,298	3,015	3,163	2,836	5,228

Table S2. Summary of SUMI identified in the human and chimp samples. The number of SUMIs identified in each individual is shown in the row labeled "SUMI" (columns 1-4). The number of those SUMIs that do not overlap a CpG island (CGI) is shown in the row labeled "not CGI". The last column of each table shows the number of SUMIs and the number of SUMIs not overlapping a CpG island identified in at least one individual of the given species.

	Term Category	Term Name	P-Value	FDR Q-val	Fold Enrichment	% regions with term
SUMIs with differential methylation	Cellular Component	plasma membrane	2.7E-10	2.3E-07	1.54	36
		plasma membrane part	1.2E-06	4.9E-04	1.56	23
		membrane	3.0E-06	8.4E-04	1.23	57
		membrane part	1.2E-05	2.5E-03	1.25	47
Germline methylated SUMIs	Molecular Function	metal ion binding	2.3E-06	6.4E-03	1.30	42
		ion binding	5.1E-06	7.1E-03	1.29	42
	Biological Process	biological regulation	4.7E-07	2.5E-03	1.19	68
		regulation of biological process	3.1E-06	8.2E-03	1.18	65
		regulation of cellular process	5.3E-06	9.2E-03	1.18	63
Germline unmethylated SUMIs			no association detected			

Table S3. GO term analysis of various SUMI types. GO term analysis was carried out using GREAT on 458 SUMIs with differential methylation between human and chimp, 493 SUMIs with evidence of germline methylation conserved in human, chimp and orang, and 775 SUMIs with evidence of germline lack of methylation conserved in human, chimp and orang. Enrichment was calculated against a background set of 11,718 orthologous human, chimp and orang SUMIs. P-values and Q-values were estimated by GREAT using the hypergeometric test.

References

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**(3): 523-536.

Singer M, Boffelli D, Dhahbi J, Schoenhuth A, Schroth GP, Martin DIK, Pachter L. 2010. MetMap Enables Genome-Scale Methyltyping for Determining Methylation States in Populations. *PLoS Comput Biol* **6**(8): e1000888.