Figure S1  The density distribution of cytosine modification levels. The β-values of 290,577 CpG sites on autosomes and sex chromosomes across 144 samples are shown. The β-values were quantile normalized and adjusted for batch effects.
Figure S2  Co-modification of CpG pairs in the YRI samples. Spearman’s $\rho$ values of CpG pairs < 5 kb apart were calculated. CpG pairs in which both cytosines fell in the upstream flanking region, 5’ UTR, gene body, 3’ UTR, and downstream flanking region of a particular gene were grouped by the distance between cytosines within a pair. Bars represent the medians of the signed $\rho^2$ across the CpG pairs for each distance group (in bp). Lines represent the intraquantile ranges of the signed $\rho^2$. 
Figure S3  Prominent variables detected by principal components analysis. The first two principal components for the M-values of the 290,577 CpG sites after quantile normalization are plotted, colored by bisulfite conversion batch (A) and array hybridization batch (B). The array hybridization batch appears to be the prominent variable. (C) The M-values of the 290,577 CpG sites were quantile normalized and adjusted for batch effects using COMBAT. Gender is the prominent variable. (D) The M-values of the 283,540 autosomal CpG sites after quantile normalization and batch correction. Population identity is the prominent variable.
Figure S4  The differential CpG sites between the CEU and YRI samples are unlikely biased by EBV transformation or intrinsic growth rate. For the 36,597 differential CpG sites, their M-values were regressed using a linear model: M-value ~ population + gender + EBV + error, or M-value ~ population + gender + iGR + error. The $-\log_{10} P$-values for EBV (A) or iGR (B) effect are plotted against the $-\log_{10} P$-values for the population effect. Among the 36,597 differential cytosines, only 146 CpGs showed EBV effect, and 246 CpGs showed iGR effect more significant than the population effect. iGR: intrinsic growth rate.
Figure S5  The 118 unrelated samples (60 CEU + 58 YRI) and 16,651 CpG sites that overlapped between this study and GSE27146 (FRASER et al. 2012) were re-processed by the same procedure. Each dataset was then re-analyzed with the linear model: cytosine modification level ~ population + gender + error. The regression coefficients of the two datasets are shown as scatter plot. The x-axis: data from FRASER et al. 2012, y-axis: data from this study.
Figure S6  Principal components analysis on cytosine modification profiles of pooled CEU, YRI and Dutch samples. Upper panel: 133 samples in this study pooled with 88 healthy individuals of Dutch descent on the 27K array (GSE41037, Plate C); Lower panel: samples in this study pooled with 33 healthy individuals of Dutch descent on the 450K array (GSE41169). The cytosine modification levels were quantile normalized and adjusted by batch. The clustering patterns were inconsistent with the difference of tissue types (whole blood vs. LCL) or cell cultures (primary vs. cultured).

(A) and (D): The principal components of all CpG sites, colored by batch (1: Dutch study; 2, 3 and 4: this study).

(B) and (E): The principal components of all CpG sites, colored by population identity (1: YRI; 2:CEU; 3:Dutch).

(C) and (F): The principal components of differential CpG sites between the CEU and YRI samples, colored by population identity (1:YRI; 2:CEU; 3:Dutch).
Figure S7  mQTLs are highly enriched within 100 kb of the target CpG sites. The proportions of significant mQTLs (detected in the 1 mb regions) for the 36,597 differential CpG sites are shown. X-axis represents the distance between mQTLs and the target CpG sites. At 5% FDR, a total of 23,924 modification-SNP associations (1,354 CpGs) were detected in the CEU samples. 17,643 modification-SNP associations (1,918 CpGs) were detected in the YRI samples.
Figure S8  Population specificity of mQTLs. The SNP association \( r^2 \) in the CEU are plotted against those in the YRI, for mQTLs detected in either the CEU or YRI samples at 5% FDR and with MAF > 0.1 in each population.
Figure S9  Common mQTLs across the CEU and YRI samples are enriched in SNPs with higher Fst values. The sampling distribution of the proportions of random SNPs with Fst values >0.05 (left panel), >0.10 (middle panel), or >0.15 (right panel) was obtained by 10,000 random draws of 5,237 SNPs controlled by the MAF distribution of the 5,237 common mQTLs shared between the CEU and YRI samples. The MAF distribution of common mQTLs was based on the CEU samples. Random SNPs and Fst values were based on the CEU and YRI data in the dbSNP v135 database. The actual proportion of the SNPs with Fst > given cutoff among the 5,237 mQTLs is shown as a solid circle. Common mQTLs refer to the mQTLs obtained from 100 kb local regions without linkage disequilibrium pruning.
Figure S10  Bisulfite sequencing and RT-PCR. Bisulfite sequencing of 4 selected population-specific CpG probes in at least 20 random samples confirms cytosine modification levels measured by the 450K array (upper 8 panels). RT-PCR confirms expression levels measured by the Affymetrix Human Exon array (lower 8 panels). (A) Probe cg19269039, located in the first exon of DENND2D (encoding DENN/MADD domain containing 2D). (B) Probe cg06959103, located within the promoter region of the TBX21 (encoding T-box 21). (C) Probe cg19810536, located in the promoter region of CR2 (encoding complement component receptor 2). (D) Probe cg02124892, located in the 5′-UTR of LIPH (encoding lipase, member H).
**Figure S11**  Potential importance of the CpG sites adjacent to the target CpG probes. Bisulfite sequencing allows for the measurement of CpG sites adjacent to the target CpGs interrogated by the 450K array. Bisulfite sequencing of LIPH (encoding lipase, member H) reveals that adjacent CpGs could be more differentially modified between populations than the target CpGs by the array. Each row represents a different sample. Each pie chart represents a CpG site, and the proportion of modified cytosines at each locus is quantified by the amount of circle filled in. The average percentage of modified cytosines at each locus is denoted above each column. The CpG measured by the 450K array probe is boxed in red. The CpGs adjacent to the boxed CpG are more variable. Particularly, the third CpG (marked by *) is significantly differentially modified between the CEU and YRI samples at p<0.05 by Student's t-test, suggesting that the CpG outside of the probe on the array could also be informative for CpGs of a given genomic region.
File S1

List of perfectly matching unique probes

Tables S1-S9

Table S1 Primers and probes used in measuring relative EBV copy numbers.
Table S2 Relative EBV copy numbers of the HapMap samples.
Table S3 Statistics of the 450K array probes.
Table S4 Primers used in bisulfite sequencing.
Table S5 Differentially modified cytosines between the CEU and YRI samples (YRI as reference).
Table S6 mQTL detected at 5% FDR within CEU and/or YRI.
Table S7 Population-specific cytosine modification correlated with differential gene expression between the CEU and YRI samples.
Table S8 m-eQTLs underlying modification-expression correlations, for which both modification and expression levels differed between the CEU and YRI samples.
Table S9 The overlap of mQTL detected in this study with the mQTL reported by previous studies.