Supplemental Information

Supplemental Figures

Fig. S1. Comparison of presumable chronological age and DNA methylation age inferred with Horvath and Hannum methods between KhoeSan and non-KhoeSan groups.

Fig. S2. Heatmap and clustering based on novel SNPs in KhoeSan. The arrow points to an outlier KhoeSan sample clustered together with non-KhoeSan samples, the individual has admixed origin and genetics.

Fig. S3. Principal component analysis plots (PC1 and PC2) for all normalized and filtered probes.

Fig. S4. Schema for extracting the sources of variance associated with batch effects from PCA of the non-KhoeSan group. The values for selected principal components from the non-KhoeSan were predicted for KhoeSan samples and their association with sex, sample group and array variables assessed with Kruskal-Wallis test.

Fig. S5. Addressing P-value inflation in differential methylation analysis.

Fig. S6. Strip chart of methylation Beta-values for the top 133 differentially methylated probes

Fig. S7. Between-group mean difference in KhoeSan vs non-KhoeSan plotted against variance in KhoeSan for top 133 probes.

Fig. S8. Heatmap and clustering of top 30 differentially methylated CpG in KhoeSan and non-KhoeSan samples. The heatmap color scale indicates methylation beta-values. Rows represent probes and columns are samples. An additional four annotation tracks for each sample include (top to bottom): sample group, sex, array, and age. Note that the normalized methylation values were not modified to remove unwanted variance, however such factors were included in the differential methylation regression model. Methylation in 30 probes is sufficient to perfectly separate samples in the KhoeSan group from the non-KhoeSan group. The four probes for SLC39A4 gene are shown on the bottom of the plot.

Fig. S9. Strip chart for methylation of CpG probes identified in Fagny *et al*. study.

Fig. S10. Correlogram for top ten principal components for all probes in a complete dataset against Illumina array quality metrics calculated using control probes.

Fig. S11. Venn diagram comparing probes identified by U-test (400 CpG - orange), not corrected for other sources of variance, such as blood composition and batch effects; and linear regression model (limma) with possible confounders included in the model (816 CpG in black; |deltaBeta| > 0.05: 133 CpG in green).

Supplemental Tables

TableS1-novelSNP.xlsx	Novel SNPs found with MethylToSNP
TableS2-DMP.xlsx	816 differentially methylated probes
TableS3-GO_ Enrichment.xlsx	GO enrichment for differentially methylated probes
TableS4-DMR.xlsx	Differentially methylated regions
TableS5-QC.xlsx	Various Illumina QC metrics
TableS6-samples.xlsx	Sample annotation and inferred leukocyte composition