Supplementary Material

Figure S1. Flowchart for Illumina HumanMethylation 450K data handling

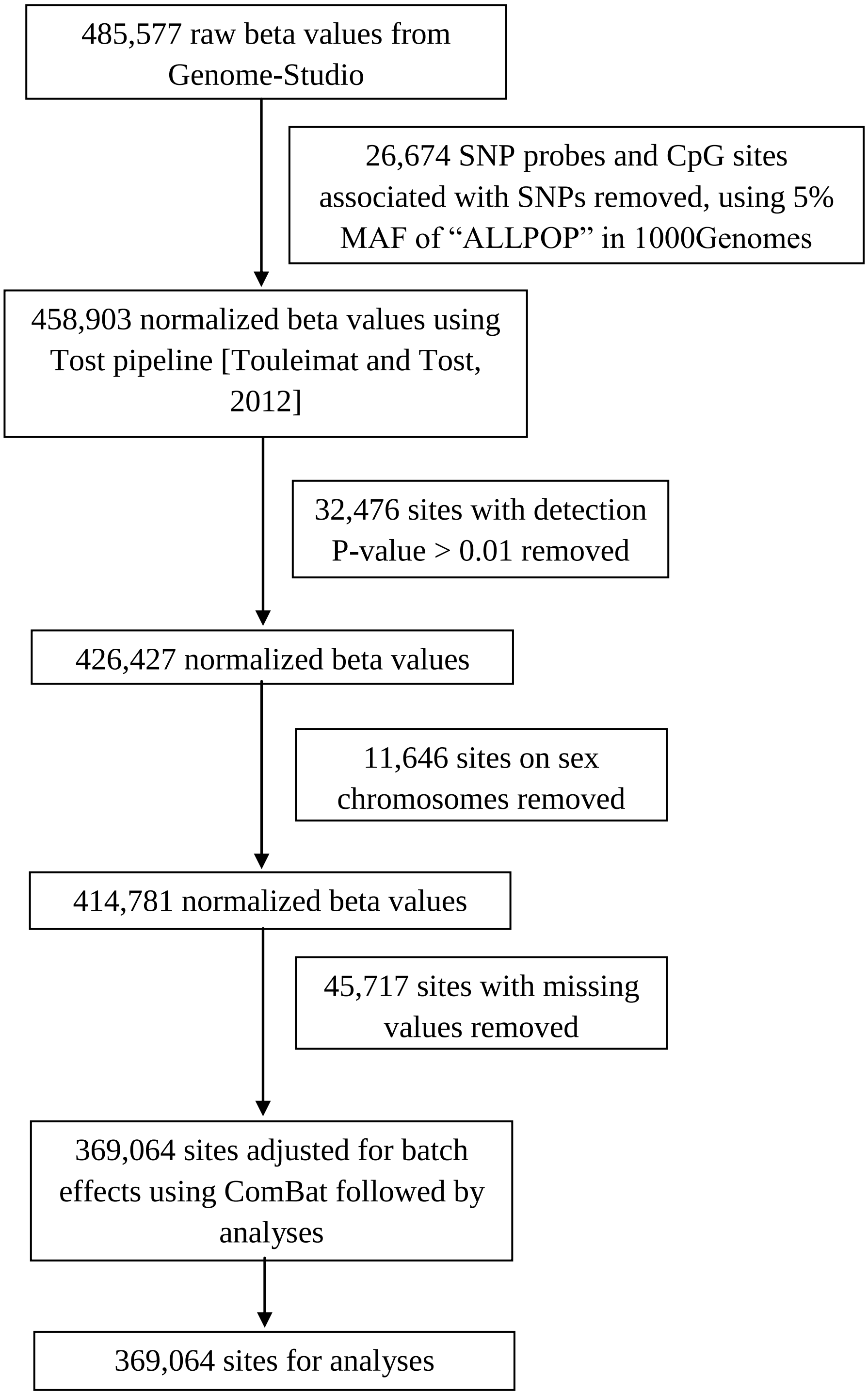

Figure S2. Density plots of chip effect and p-values with (red) and without (black) ComBat adjustment

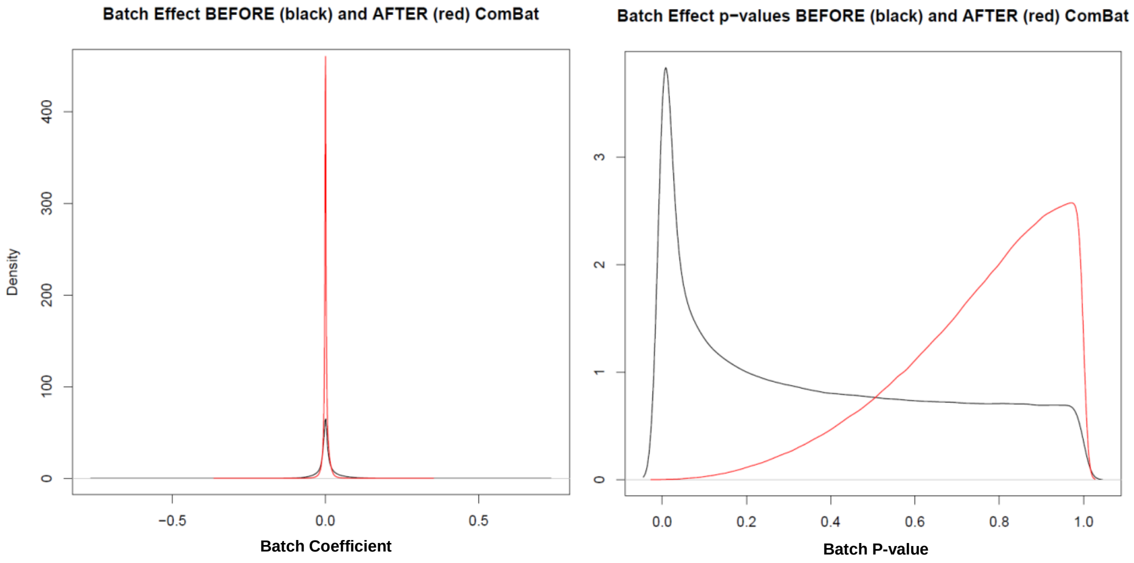

Table S1. Primers used for pyrosequencing

Infinium ID	Gene	Forward primer	Reverse primer (biotin labeled)	Sequencing primer	Sequencing entry
cg24883732	*TCEB3B*	GTAGTTTTGTTTTTGGGGTTG	CTCCAAACTATAACTCCACATCC	GAAGTTTGGGGTGGT	C/tGGTTTTC/tGTTGTTC/tGGTC/tGTTAG
cg15059639	*MYO3B*	TTGTTTTGGAGTAGAGAGTGTATAGATATT	ACAATCTCCCAATCTAACTTCTACC	ATTAGTTTGAGAGTT	C/tGATTTTGTTTTTTGTTAAGAAATGTTTC/tGTTT
cg00218770	*CDH4*	GGAAAGAGAAAAGGTTGGTGATTA	TCCAAATAACTAACAATCCTAAATC	TATGGAATTTGAGTA	C/tGTTTTATTGAATTATATC/tGTTAG
cg03333116	*RHBDF1*	GTAGGTAGGGAAGRGGGGTAGAG	CCCAATAAATAATAAATTCCCAATC	TTAGGAATAGTTGGG	C/tGAGTATC/tGGTATTTATTTTTGTGGGC/tGATGGAAG
cg10090568	*ESRRG*	AAGTGGGTAGGATTGTAGATAAATA	TAAATCCTCAAACTTTCCTATCATT	ATTTTTATGTTATTA	C/tGTGATTTTATA
cg01667892	*PRDM9*	GTTTTTTGATTGTTGAATGG	TTCTCTCAAAATCCCTCAAAAAATA	TTGATTGTTGAATGG	C/tGGGTTC/tGTTTTGGTTTTC/tGTGTTG

Table S2. Average pairwise correlation across multiple loci for each CpG site

CpG site	Gene	Location	Average pairwise correlation[a]	R^2 between duplicates
cg24883732	*TCEB3B*	Island	0.97	0.97
cg15059639	*MYO3B*	-	0.99	0.99
cg00218770	*CDH4*	N_Shore	0.44	0.91
cg03333116	*RHBDF1*	Island	0.10	0.75
cg10090568	*ESRRG*	-	-	0.97
cg01667892	*PRDM9*	-	0.89	0.97

[a] Spearman correlation was used to calculate pairwise correlation between loci.

N_Shore, North shore which are regions flanking island

Table S3. Top 20 CpG sites with the largest M-value difference in New Cases compared to Persistent Controls

CpG site	Gene	New Cases M-value difference[a]	Persistent Controls M-value difference[a]	Lower 95% CI	Upper 95% CI	*P*[b]
cg24883732	*TCEB3B*	1.23	-0.42	-2.99	-0.31	0.02
cg15059639	*MYO3B*	-1.04	0.62	0.69	2.63	0.003
cg03065888	*CYC1*	0.93	-0.25	-1.94	-0.41	0.005
cg19296405	*KIAA1683*	-1.08	0.12	0.46	1.92	0.003
cg00218770	*CDH4*	1.07	-0.22	-2.38	-0.19	0.02
cg03333116	*RHBDF1*	1.05	-0.66	-2.45	-0.97	0.0002
cg06352616	*CCDC154*	0.79	-0.19	-2.27	0.31	0.12
cg03371700	*JAKMIP3*	0.94	-0.01	-1.8	-0.1	0.03
cg11511175	*AGAP2*	0.72	-0.36	-2.2	0.03	0.06
cg17900076	*LOC642597*	-0.76	0.46	0.3	2.13	0.01
cg24405174	*PL-5283*	0.83	0.12	-1.38	-0.04	0.04
cg00028056	*CHPF*	0.87	-0.33	-2.03	-0.37	0.01
cg02916425	*C2orf81*	-0.94	0.24	0.03	2.34	0.05
cg10090568	*ESRRG*	-0.7	0.11	0.33	1.31	0.002
cg27370573	*PPAP2C*	0.99	-0.01	-1.89	-0.1	0.03
cg22217449	*PCNT*	0.81	-0.44	-2.3	-0.19	0.03
cg07664579	*SLC6A3*	0.94	0.02	-1.62	-0.23	0.01
cg01667892	*PRDM9*	-0.76	-0.08	0.09	1.27	0.03
cg25148456	*MAP3K1*	0.8	0.09	-1.35	-0.07	0.03
cg02749948	*TNXB*	-0.86	0.25	0.11	2.1	0.03

[a] M-value differences were calculated as M-value at follow-up (2009–2011) subtracted by M-value at baseline (2001–2003).

[b] P-values were obtained from two-sample t-tests for comparisons of M-value differences between New Cases and Persistent Controls.