Supplementary Methods

Quality Control (QC)

Removing batch effects

As shown in Supplementary Figure 2, there are differences in the mean methylation level between individuals that may be a result of the position in which their DNA was plated on the chip. In order to remove batch effects and effects due to samples being in different positions on the plates, a linear regression model was fitted with the scores for each probe as the dependent variable and chip, column and row as the independent variables as factors. The residuals were used for further analysis.

Removal of probes with excessive missingness

In order to remove probes that showed high levels of missingness, we first estimated the overall missingness rate to be 0.4% (total of 127357 missing values across 485204 probes in 56 individuals). We calculated that 5.1% of probes had 2 or more missing values, much greater than than the 2.1% expected if missingness was by chance based on binomial probability. Therefore, 24,744 with 2 or more missing values were removed, leaving 472,629 probes.

The next QC steps probes were grouped according to their relationship to CpG islands because average beta values are associated with these groupings. CpG island are defined as regions rich in CpG sites close to the promoter regions of the genes, shores are defined as regions flanking CpG islands, and shelves as regions flanking shores and non-CpG (NCpG) as isolated CpGs in the genome. Probe annotations were provided by Illumina (Illumina, San Diego, CA, USA).

Comparison with results from HM 27 BeadChip

The HM 450 BeadChip includes 90% of the CpG sites contained on the HM 27 BeadChip. Data from the five female MZT pairs previously epityped with the HM 27 BeadChip were matched with the HM

450 data. Out of 472,629 (HM 450 BeadChip) and 27,578 (HM 27BeadChip), 25,394 probes could be matched. Correlations for each of the samples included in the HM 27 BeadChip and the HM 450 BeadChip showed very good agreement between the data from the two arrays (correlation of β values across all samples was >0.98), demonstrating the reliability of the new array.

Removal of probes with high degree of discordance between technical replicates

Our experiment was designed to investigate differences in methylation patterns between genetically identical individuals. Our second QC (QC2) step sought to exclude probes that showed non-random large differences in transformed β values between technical replicate samples. The replicates were typed either on different chips or in non-adjacent wells, whereas the twin pairs were typed on the same chip in neighbouring wells, so QC undertaken on the replicated samples should be conservative. For each of the 6 replicated samples, we calculated the difference between β values for each probe and then the mean and SD of these differences across probes. We labelled any difference >±3sd as an "outlier". Out of 143,391 probes in CpG Islands there were 18,142 (2%) outliers. We calculated that 3.2% of probes had two or more outliers. Under binomial theory, if outlier status was random across probes we would expect only 0.5% of probes to harbour 2 or more outliers. A similar excess of probes showing differences between replicate pairs was found for probes in shelves, shores and non-CpG islands. On this basis we excluded all probes which generated absolute differences of >±3 sd in 2 or more replicate samples. The number of probes removed, by probe annotation, is shown in Supplementary Table 6. After removing probes with large amounts of missing data and high levels of discordance between replicate pairs, a total of 462,002 (95.1%) probes remained for analysis.

Testing for Methylation Differences by Sex

In addition to testing for methylation status associated with MDD, a further analysis testing for the effects of sex on methylation was performed. The probe intensity residuals for each individual were

fitted in a linear model with sex, set, family, and case-control status. Probes found on the X and Y chromosome were not included in the analysis. 452,275 probes remained.

The Q-Q plot of the observed distribution of p-values for gender differences versus the expected uniform distribution is shown in Supplementary Figure 7. The results indicate that there are significant differences in methylation status between the sexes across many probes. After applying a stringent Bonferroni correction significance threshold of 10^{-7} ($p = 0.05$ corrected for 454,000 tests), 4,863 probes reached experiment-wise significance. Supplementary Table 4 lists the most significant probes ($p < 10^{-17}$). The most significant probe was cg15083522, found in the 3'UTR region of the *LRRC27* gene. The function of this gene is not characterised. Significant differences between the sexes were observed for all probes regardless of their being in a CpG island, shore, shelf or non CpG island (results not shown).

A two-sample t-test and F-test were performed to test for differences in the means and variances respectively of males and females across all autosomal probes. The differences in the mean and variance were both significant ($p < 2.2 \times 10^{-22}$ and $p < 2.2 \times 10^{-16}$ respectively), with males having on average more methylation and higher variance of methylation.

Pathway analysis of a list of 740 genes with at least one probe significant at $p < 10^{-7}$ in Ingenuity identified the Antigen Presentation Pathway as the only significantly enriched pathway after correction for multiple testing using FDR ($p = 0.02$). The list of genes found in this pathway consists of 6 genes in the MHC region. The next most significant pathway was Cytotoxic T-Lymphocyte Mediated Apoptosis of Target Cells. A similar analysis in DAVID identified actin-binding and actin-cytoskeleton as the most significantly enriched annotation cluster ($p = 0.03$). Activation of T cells and Immunoglobulin domains were among the most significantly enriched annotations, although they did not reach the significance threshold ($p = 0.11$).

Supplementary Tables and Figures

Supplementary Table 1: Phenotypic Information (Females). (Set) Discordant and matched concordant pair (1) Age of blood sample collection. (2) Ever consumed drugs >10 times p/year

Set	Age[1]	Status	Recurrent MDD	Alcohol dependent	Drugs[2]	Smoke status	Smoke p/day	Prescribed Antidepressants
1Discordant	39	MDD	Yes	No	No	Never		Yes
1Discordant	39	No MDD		No	No	Never		
1Concordant	41	No MDD		No	No	Never		
1Concordant	41	No MDD		No	No	Ex-smoker	5-10/day	
2 Discordant	35	MDD	Yes	No	No	Never		Yes
2 Discordant	35	No MDD		No	No	Never		
2 Concordant	34	No MDD		No	No	Never		
2 Concordant	34	No MDD		No	No	Never		
3 Discordant	56	MDD	Yes	No	No	Never		Yes
3 Discordant	56	No MDD		No	No	Never		
3 Concordant	59	No MDD		Yes	No	Never		
3 Concordant	59	No MDD		No	No	Never		
4 Discordant	39	MDD	Yes	No	No	Never		
4 Discordant	39	No MDD		No	No	Never		
4 Concordant	32	No MDD		No	No	Never		
4 Concordant	32	No MDD		No	No	Never		
5 Discordant	58	MDD	Yes	No	No	Never		Yes
5 Discordant	58	No MDD		No	No	Never		
5 Concordant	63	No MDD		No	No	Never		
5 Concordant	63	No MDD		No	No	Never		
6 Discordant	52	MDD	Yes	No	No	Ex-smoker	1-4/day	
6 Discordant	52	No MDD		No	No	Never		
6 Concordant	51	No MDD		No	No	Never		
6 Concordant	51	No MDD		No	No	Never		

Supplementary Table 2: Phenotypic Information (Males). (Set) Discordant and matched concordant pair (1) Age of blood sample collection. (2) Ever consumed drugs >10 times p/year. (?) Information not known.

Set	Age[1]	Status	Recurrent MDD	Alcohol dependent	Drugs[2]	Smoke status	Smoke p/day	Prescribed Antidepressants
7 Discordant	61	MDD	Yes	Yes	No	Never		Yes
7 Discordant	61	No MDD		No	No	Never		
7 Concordant	58	No MDD		No	No	Current	?	
7 Concordant	58	No MDD		No	No	Ex-smoker	1-4/day	

8 Discordant	50	MDD	No	No	Yes	Never		
8 Discordant	50	No MDD		No	No	Never		
8 Concordant	48	No MDD		Yes	No	Ex-smoker	21-40/day	
8 Concordant	48	No MDD		No	No	Ex-smoker	11-20/day	
9 Discordant	41	MDD	Yes	Yes	Yes	Current	21-40/day	Yes
9 Discordant	41	No MDD		No	No	Current	21-40/day	
9 Concordant	46	No MDD		No	No	Ex-smoker	11-20/day	
9 Concordant	46	No MDD		No	No	Ex-smoker	21-40/day	
10 Discordant	44	MDD	No	No	?	Never		
10 Discordant	44	No MDD		No	?	Never		
10 Concordant	37	No MDD		No	?	Never		
10 Concordant	37	No MDD		No	?	Never		
11 Discordant	32	MDD	No	No	?	Never		
11 Discordant	32	No MDD		No	?	Never		
11 Concordant	36	No MDD		No	?	Never		
11 Concordant	36	No MDD		No	?	Never		
12 Discordant	31	MDD	No	Yes	No	Ex-smoker	1-4/day	
12 Discordant	31	No MDD		No	No	Current	11-20/day	
12 Concordant	35	No MDD		No	?	Never		
12 Concordant	35	No MDD		No	?	Never		

Supplementary Table 3: Sample allocation by bisulphite run.

Bisulphite run	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**
MZT pairs	4	4	4	4	3	3	2 + 2*	4*
MZT technical replicate	1	1	1	1	1	1	0	0
Technical replicate	CEPH ♀	CEPH ♀	CEPH ♀	CEPH ♀	CEPH ♀ CEPH ♂ FSK	CEPH ♀ CEPH ♂ FSK	CEPH ♀ FSK	CEPH ♀ FSK
Total (No. Samples)	10	10	10	10	10	10	10	10

Supplementary Table 4: HM 450 BeadChips. Sample allocation

	Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**
MZT pairs	5	5	5	6	3
MZT technical replicate	1	1	1	0	1
Technical replicate (FSK)	1	1	1	0	1
Total (No. Samples)	12	12	12	12	8

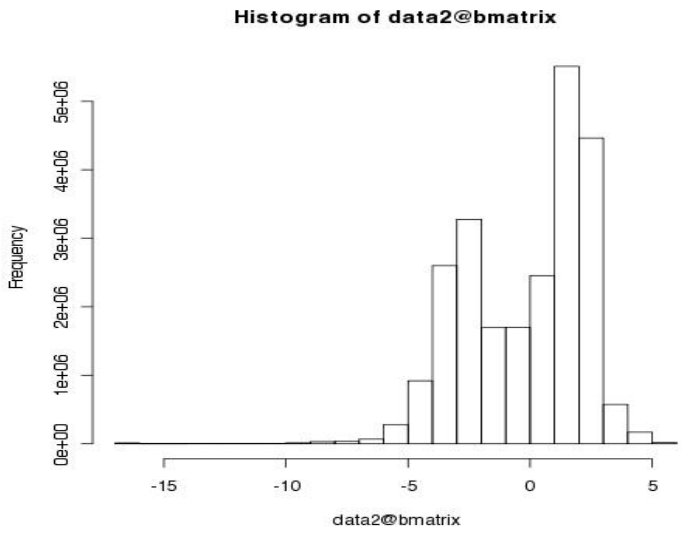
Supplementary Table 5. Results from comparing distribution of variance across different probe types in cases and controls.

Annotation	**Number of Probes**	**Wilcoxon sign-rank test p-value**	**% probes increased variance in cases**	**Binomial P**
All probes	462,001	< 2.2 x 10E-16	52.3	< 10E-16
Islands	138,738	< 2.2 x 10E-16	51.7	< 10E-16
Shelfs	45,691	< 2.2 x 10E-16	53.1	< 10E-16
Shores	107,394	2.9 x 10E-07	51.2	<10E-16
No annotation	170,178	< 2.2 x 10E-16	53.4	<10E-16

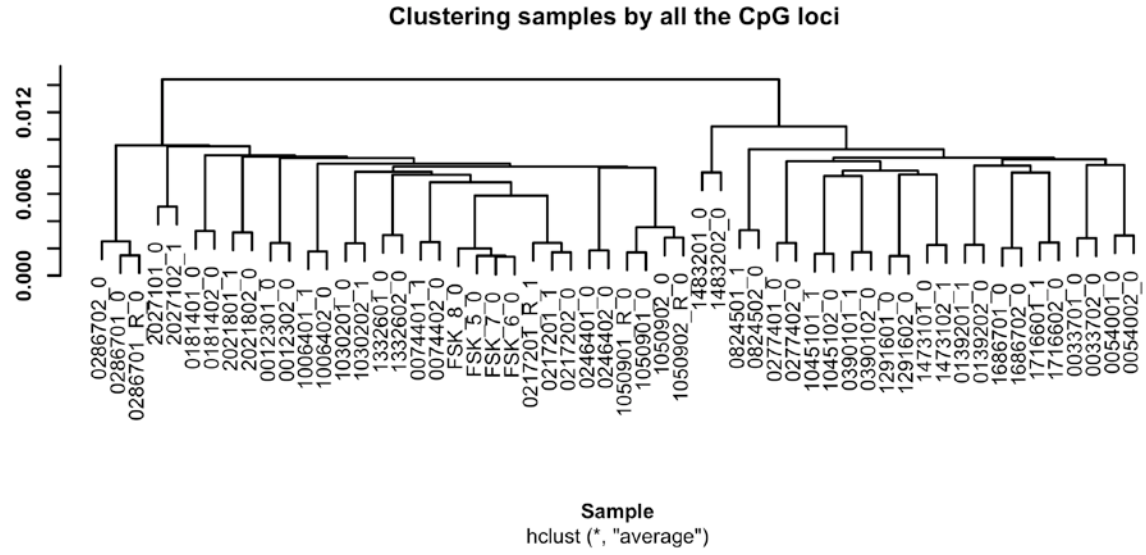
Supplementary Table 6. Autosomal probes most significantly differentiated by sex.

ILMNID	CHR	MAPINFO	UCSC_REFGENE NAME	UCSC_REFGENE_GROUP	RELATION_TO_UCSC_CPG_ISLAND	Pval
cg15083522	10	134188873	LRRC27	3'UTR	NA	2.97E-21
cg27308738	10	105357975	SH3PXD2A	3'UTR	N_Shelf	1.17E-20
cg27079096	11	4389638	OR52B4	TSS200	NA	9.91E-20
cg01188578	2	26464058	HADHA	Body	N_Shelf	1.51E-19
cg15602423	6	32552095	HLA-DRB1	Body	Island	1.55E-19
cg03020684	15	71532066	THSD4	Body	NA	1.57E-19
cg18709904	14	50474530	C14orf182	TSS1500	NA	2.68E-19
cg11606607	17	78264297	RNF213	Body	NA	5.64E-19
cg00325917	1	169671116	SELL	Body	NA	9.55E-19
cg09931872	10	49909285	WDFY4	5'UTR	NA	2.18E-18
cg01943931	14	73373205	NA	NA	NA	3.40E-18
cg20022541	14	94385395	FAM181A;C14orf86	5'UTR;Body	NA	4.93E-18
cg09351263	16	85864047	NA	NA	S_Shore	7.07E-18
cg07903626	16	66098650	NA	NA	NA	8.78E-18

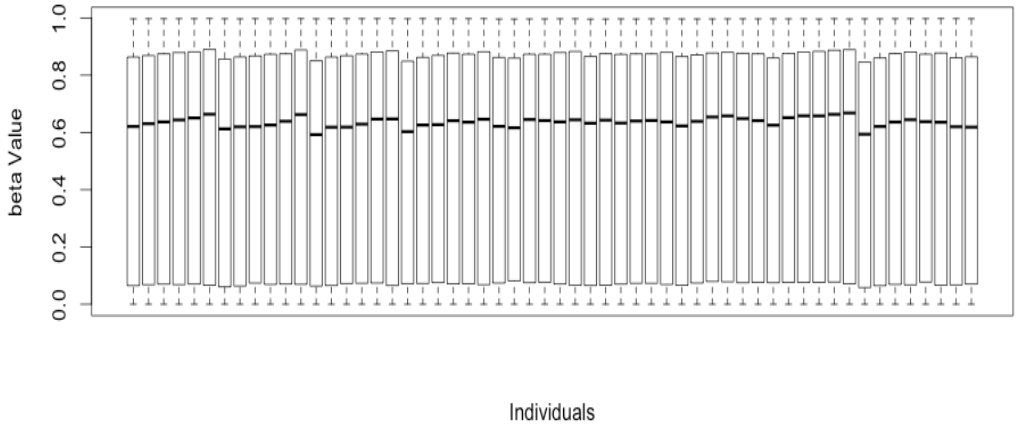
Supplementary Figure 1. Histogram of logit transformed beta values



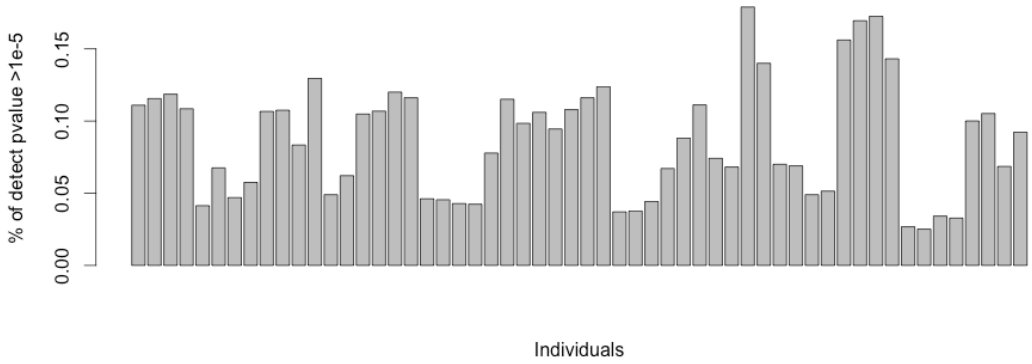
Supplementary Figure 2. Unsupervised clustering of samples based on all CpG sites



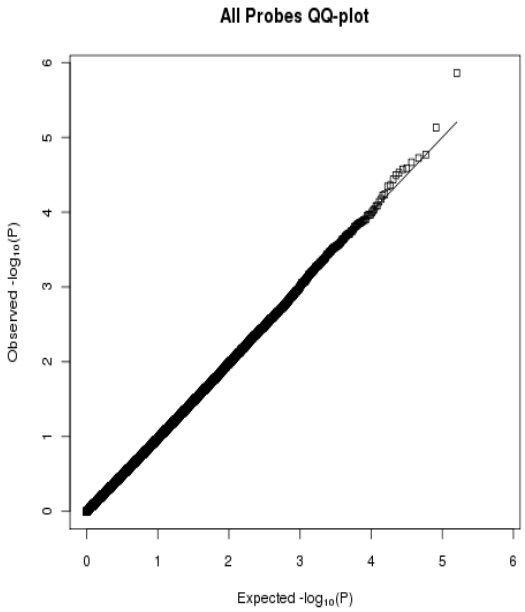
Supplementary Figure 3. Boxplots of raw beta values across all samples



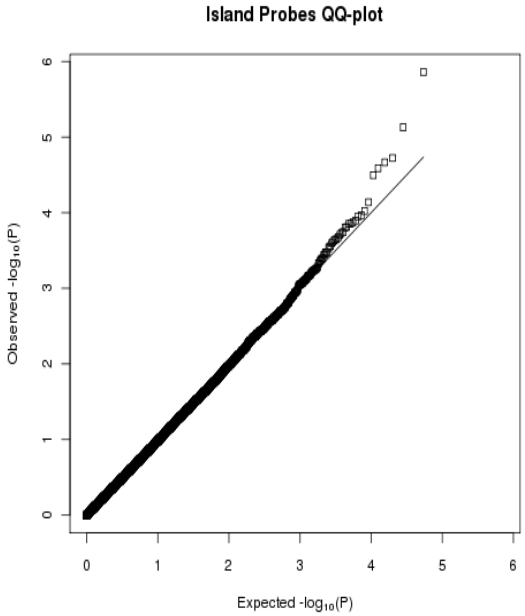
Supplementary Figure 4. Histogram of %probes by individual with detection p > 10^{-5}



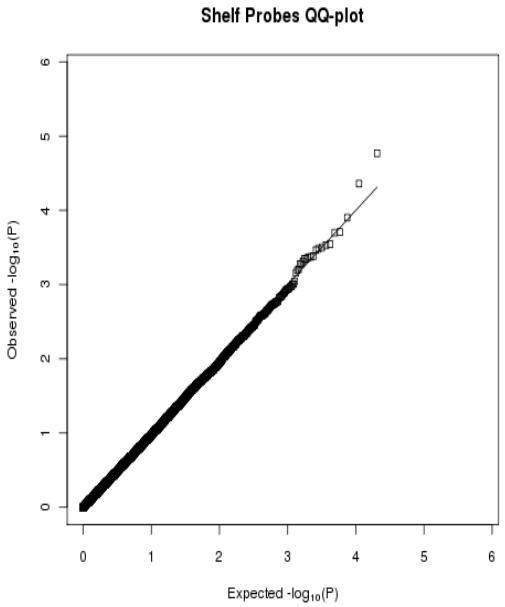
Supplementary Figure 5a. Q-Q plot for MDD analysis of all probes



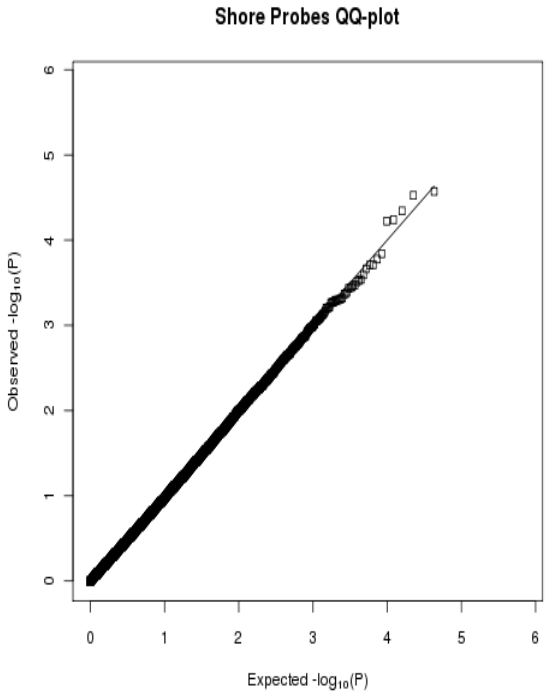
Supplementary Figure 5b Q-Q plot for all probes in CpG islands



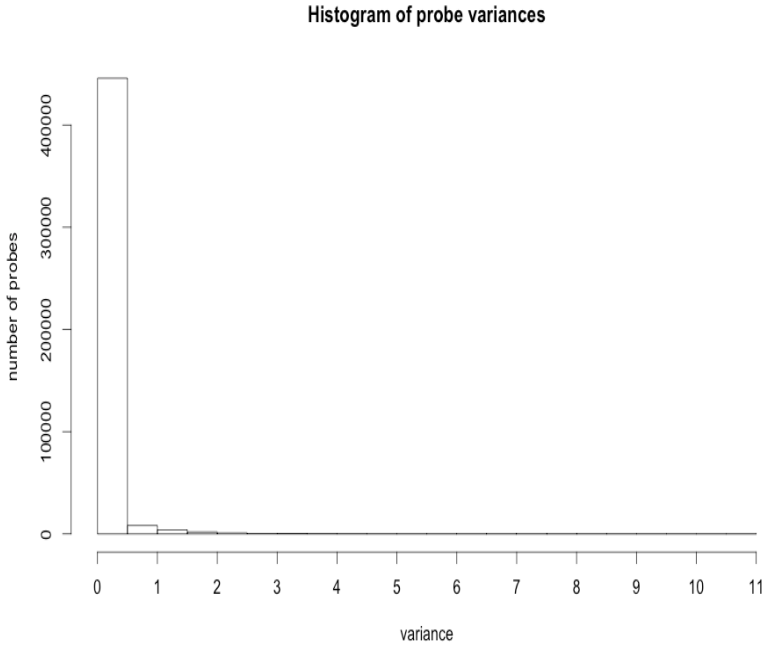
Supplementary Figure 5c. Q-Q plot of all probes in shelfs



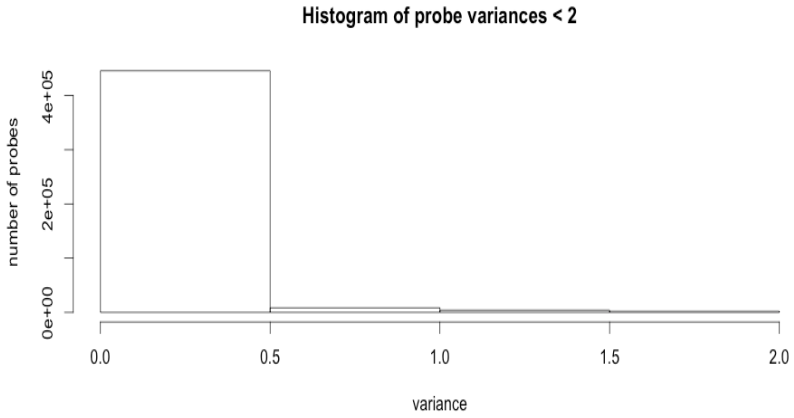
Supplementary Figure 5d. QQ plot of all probes in shores



Supplementary Figure 6a. Histogram of probe variances for all probes



Supplementary Figure 6b. Histogram of probe variances for probes with variance less than 2.



Supplementary Figure 7. Q-Q plot of analysis of methylation status by sex.