

A Full Bayesian Partition Model for Identifying Differentially Methylated Loci from Single Nucleotide Resolution Sequencing Data

Henan Wang, Chong He, Garima Kushwaha, Dong Xu, Jing Qiu

November 6, 2015

To make inference about the membership of each loci along the whole genome, the derivation for the posterior distribution of \mathbf{I} can be given by

$$\begin{aligned}
 p(\mathbf{I}|\mathbf{C}, \mathbf{M}) &= \frac{p(\mathbf{C}, \mathbf{M}|\mathbf{I})p(\mathbf{I})}{\int p(\mathbf{C}, \mathbf{M}|\mathbf{I})p(\mathbf{I})d\mathbf{I}} \\
 &= \frac{p(\mathbf{C}|\mathbf{I})p(\mathbf{M}|\mathbf{C}, \mathbf{I})p(\mathbf{I})}{\int p(\mathbf{C}|\mathbf{I})p(\mathbf{M}|\mathbf{C}, \mathbf{I})p(\mathbf{I})d\mathbf{I}} \\
 &= \frac{p(\mathbf{M}|\mathbf{C}, \mathbf{I})P(\mathbf{I})}{\int p(\mathbf{M}|\mathbf{C}, \mathbf{I})p(\mathbf{I})d\mathbf{I}} \quad (\text{since } \mathbf{C} \text{ and } \mathbf{I} \text{ are independent, } p(\mathbf{C}|\mathbf{I}) = p(\mathbf{C})) \\
 &= \frac{p(\mathbf{M}_0|\mathbf{C}, \mathbf{I})p(\mathbf{M}_1|\mathbf{C}, \mathbf{I})p(\mathbf{M}_2|\mathbf{C}, \mathbf{I})p(\mathbf{I})}{\int p(\mathbf{M}|\mathbf{C}, \mathbf{I})p(\mathbf{I})d\mathbf{I}} \quad (\text{assume } \mathbf{M}_0, \mathbf{M}_1 \text{ and } \mathbf{M}_2 \text{ are independent}) \\
 &= \frac{p(\mathbf{M}_0|\mathbf{C}_0, \mathbf{I})p(\mathbf{M}_1|\mathbf{C}_1, \mathbf{I})p(\mathbf{M}_2|\mathbf{C}_2, \mathbf{I})p(\mathbf{I})}{\int p(\mathbf{M}|\mathbf{C}, \mathbf{I})p(\mathbf{I})d\mathbf{I}} \\
 &\propto p(\mathbf{M}_0|\mathbf{C}_0, \mathbf{I})p(\mathbf{M}_1|\mathbf{C}_1, \mathbf{I})p(\mathbf{M}_2|\mathbf{C}_2, \mathbf{I})p(\mathbf{I})
 \end{aligned}$$

Tables

Table S1 – The proportion of hypo- and hyper-methylated loci in the real data at different γ values.

	Hypo-methylated	Hyper-methylated
$\gamma = 0.01$	43.28%	29.44%
$\gamma = 0.05$	20.58%	16.79%
$\gamma = 0.1$	12.12%	10.28%
$\gamma = 0.15$	8.28%	6.77%
$\gamma = 0.2$	5.9%	4.6%

Table S2 – Comparison in identification of the four procedures (logistic regression, DSS, z -test and the proposed Bayesian) at different γ values for the simulation studies without subject effect. Results are averaged over 100 replications of 20,000 CpG loci with two samples under each condition.

True I	I from logistic			I from DSS			I from z test			I from Bayesian			
	0	1	2	0	1	2	0	1	2	0	1	2	
$\gamma = 0.01$	0	5231.04	7.73	6.13	5244.83	0	0.07	5228.42	10.64	5.84	5210.42	21.17	13.31
	1	6640.29	1831.91	0.94	7835.93	637.19	0.02	6668.5	1804.05	0.59	6270.51	2199.77	2.86
	2	4933.68	0.69	1347.59	5683.67	0	598.29	5094.69	0.56	1186.71	4725.64	4.15	1552.17
$\gamma = 0.05$	0	11718	28.83	22.22	11768.62	0.17	0.26	11715.12	34.17	19.76	11705.7	37.2	26.15
	1	2743.55	1769.78	0.15	3795.98	717.5	0	2846.05	1667.36	0.07	2594.19	1918.86	0.43
	2	2309.26	0.22	1407.99	3058.6	0	658.87	2480.87	0.15	1236.45	2262.63	0.68	1454.16
$\gamma = 0.1$	0	14851.26	33.19	24.97	14908.74	0.27	0.41	14846.17	40.58	22.67	14844.66	37.75	27.01
	1	1273.35	1428.39	0	1970.2	785.54	0	1374.93	1380.81	0	1135.53	1620.18	0.03
	2	1069.24	0.01	1265.59	1626.9	0	707.94	1188.76	0	1146.08	1015.18	0.02	1319.64
$\gamma = 0.15$	0	16489.19	27.29	21.85	16537.6	0.29	0.44	16482.05	36.29	19.99	16483.35	31.41	23.57
	1	643.51	1260.08	0	1100.23	803.36	0	709.99	1193.6	0	523.88	1370.71	0
	2	501.82	0	1056.26	857.36	0	700.72	568.3	0	989.78	449.84	0	1108.24
$\gamma = 0.2$	0	17532.55	20.88	17.66	17570.3	0.3	0.49	17523.16	29.63	18.3	17527.48	25.41	18.2
	1	313.79	1055.4	0	598.59	770.6	0	352.99	1016.2	0	239.25	1129.94	0
	2	228.03	0	831.69	433.44	0	626.28	260.05	0	799.67	194.67	0	865.05

Table S3 – The relative power improvement of the proposed Bayesian method compared to three other methods (logistic regression, z -test and DSS) in terms of TPR, TPR_hypo and TPR_hyper at different γ values for the simulation studies without subject effect. Results are averaged over 100 replications of 20,000 CpG loci with two samples under each condition.

		logistic	z test	DSS
$\gamma = 0.01$	TPR	18.16%	25.64%	204.2%
	TPR_hypo	20.08%	21.94%	245.2%
	TPR_hyper	15.18%	30.8%	159.4%
$\gamma = 0.05$	TPR	6.17%	16.19%	145.1%
	TPR_hypo	8.42%	15.08%	167.4%
	TPR_hyper	3.28%	17.61%	120.7%
$\gamma = 0.1$	TPR	6.98%	16.34%	96.85%
	TPR_hypo	9.29%	17.34%	106.3%
	TPR_hyper	4.27%	15.14%	86.41%
$\gamma = 0.15$	TPR	7.02%	13.54%	64.82%
	TPR_hypo	8.78%	14.84%	70.62%
	TPR_hyper	4.92%	11.97%	58.16%
$\gamma = 0.2$	TPR	5.72%	9.86%	42.82%
	TPR_hypo	7.06%	11.19%	46.63%
	TPR_hyper	4.01%	8.18%	38.13%

Table S4 – Similar to Table S2, but this is the result for the simulation studies with subject effect.

	True I	I from logistic			I from DSS			I from z test			I from Bayesian		
		0	1	2	0	1	2	0	1	2	0	1	2
$\gamma = 0.01$	0	5237.13	4.67	3.1	5244.9	0	0	5235.63	6	3.27	5219.4	16.04	9.46
	1	6822.06	1650.74	0.34	8331.7	141.44	0	6847.71	1625.25	0.18	6314.54	2157.04	1.56
	2	5051.41	0.24	1230.31	6171.26	0	110.7	5207.47	0.2	1074.29	4760.34	2.56	1519.06
$\gamma = 0.05$	0	11740.41	16.67	11.97	11769.05	0	0	11738.63	18.79	11.63	11721.98	28.38	18.69
	1	2886.83	1626.63	0.02	4344	169.48	0	2982.41	1531.07	0	2622.28	1891.06	0.15
	2	2418.57	0.02	1298.88	3587.44	0	130.03	2585.75	0.02	1131.7	2284.87	0.23	1432.37
$\gamma = 0.1$	0	14878.56	18.11	12.75	14909.42	0	0	14875.87	20.96	12.59	14863.17	26.46	19.79
	1	1372.06	1383.68	0	2565.07	190.67	0	1477.79	1277.95	0	1147.56	1608.17	0.01
	2	1146.35	0	1188.49	2185.25	0	149.59	1277.16	0	1057.68	1018.64	0	1316.2
$\gamma = 0.15$	0	16511.59	15.88	10.86	16538.33	0	0	16507.75	19.81	10.77	16499.03	23.17	16.13
	1	702.6	1200.99	0	1690.3	213.29	0	779.57	1124.02	0	532.09	1371.5	0
	2	548.95	0	1009.13	1399.71	0	158.37	626.26	0	931.82	451.74	0	1106.34
$\gamma = 0.2$	0	17549.52	12.7	8.87	17571.07	0	0.02	17545.16	16.55	9.38	17539.91	18.64	12.54
	1	347.95	1021.24	0	1148.5	220.69	0	393.09	976.1	0	234.5	1134.69	0
	2	248.11	0	811.61	897.69	0	162.03	286.34	0	773.38	189.67	0	870.05

Table S5 – Similar to Table S3, but this relative power improvement is for the simulation studies with subject effect.

		logistic	z test	DSS
$\gamma = 0.01$	TPR	27.71%	36.31%	1359.59%
	TPR_hypo	30.67%	32.72%	1425.06%
	TPR_hyper	23.47%	41.4%	1272.23%
$\gamma = 0.05$	TPR	13.61%	24.82%	1009.75%
	TPR_hypo	16.26%	23.51%	1015.8%
	TPR_hyper	10.28%	26.57%	1001.57%
$\gamma = 0.1$	TPR	13.69%	25.21%	759.45%
	TPR_hypo	16.22%	25.84%	743.43%
	TPR_hyper	10.75%	24.44%	779.87%
$\gamma = 0.15$	TPR	12.11%	20.53%	566.7%
	TPR_hypo	14.2%	22.02%	543.02%
	TPR_hyper	9.63%	18.73%	598.6%
$\gamma = 0.2$	TPR	9.38%	14.59%	423.8%
	TPR_hypo	11.11%	16.25%	414.16%
	TPR_hyper	7.2%	12.5%	436.97%

Table S6 – Numbers of hypo- and hyper-methylated loci identified by DSS and z -test method for the real data. FDR is controlled at levels 0.01, 0.02, 0.03 and 0.05 for both methods.

	hypo-methylation	hyper-methylation
DSS: FDR controlled at 0.01	4,066	2,814
DSS: FDR controlled at 0.02	4,801	3,258
DSS: FDR controlled at 0.03	5,316	3,636
DSS: FDR controlled at 0.05	6,119	4,142
z -test: FDR controlled at 0.01	20,365	13,126
z -test: FDR controlled at 0.02	24,300	15,541
z -test: FDR controlled at 0.03	27,390	17,373
z -test: FDR controlled at 0.05	32,111	20,070

Table S7 – Summary of simulation studies to study the proportions of false positives in the list of uniquely identified loci by either the proposed Bayesian method or the logistic regression method at 0.05 FDR level (the results are averaged over 100 simulation datasets with subject effect when γ is 0.2). Here “total” indicates the total number of uniquely identified loci by either method and “total with difference >0.2 ” indicates the number of loci uniquely identified by either method with methylation proportion difference greater than 0.2. Similarly, we also check situations when the methylation proportion difference is less or equal than 0.2, and less than 0.1 respectively.

	Logistic regression	The proposed Bayesian
total	115.21	28.52
proportion of false positives (FPs) among uniquely identified DML	57.42%	13.86%
total with difference >0.2	30.92(26.84%)	25.17(88.25%)
proportion of FPs among the uniquely identified DML with difference >0.2	7.44%	11.14%
proportion of FPs among the uniquely identified DML with difference ≤ 0.2	75.8%	33.73%
total with difference <0.1	45.17(39.21%)	0(0%)
proportion of FPs among the uniquely identified DML with difference <0.1	99.48%	0

Table S8 – List of genes with hyper-methylated CpGs uniquely identified by the proposed Bayesian model.

<i>AAAS</i>	<i>CENPV</i>	<i>FGF22</i>	<i>LIMD2</i>	<i>NFATC4</i>	<i>RP4-738P15.1</i>	<i>TMCC1</i>
<i>AC009473.1</i>	<i>CGB7</i>	<i>GALE</i>	<i>LINC00299</i>	<i>NRK</i>	<i>RRAGA</i>	<i>TMEM181</i>
<i>ADCK2</i>	<i>CHRM5</i>	<i>GPER1</i>	<i>LMNB2</i>	<i>ODC1</i>	<i>RTN1</i>	<i>TSPAN17</i>
<i>AFF4</i>	<i>CHRNA1</i>	<i>GPR113</i>	<i>LPCAT4</i>	<i>OR7E91P</i>	<i>RTP4</i>	<i>TUSC5</i>
<i>AP1M2</i>	<i>CKMT2-AS1</i>	<i>GRIA3</i>	<i>LUZP2</i>	<i>OTUD7A</i>	<i>SERPINB6</i>	<i>UBA52</i>
<i>ARF5</i>	<i>CLIP2</i>	<i>HIF1AN</i>	<i>MAN1C1</i>	<i>PDE11A</i>	<i>SH3BGRL3</i>	<i>UBL4B</i>
<i>ARFGEF2</i>	<i>CNKSR3</i>	<i>HIST1H2BO</i>	<i>MARCO</i>	<i>PINX1</i>	<i>SIRT6</i>	<i>URI1</i>
<i>ASPSR1</i>	<i>CREBRF</i>	<i>HMBOX1</i>	<i>MDH1B</i>	<i>PIP5KL1</i>	<i>SIX4</i>	<i>WDR89</i>
<i>ATE1</i>	<i>CST7</i>	<i>HMHA1</i>	<i>MIR326</i>	<i>PLEKHA7</i>	<i>SLC29A1</i>	<i>WNT3</i>
<i>AZI1</i>	<i>CYB5A</i>	<i>HOXA1</i>	<i>MIR3659</i>	<i>PLEKHH2</i>	<i>SLC38A11</i>	<i>YWHAZ</i>
<i>B4GALT7</i>	<i>DFFB</i>	<i>HPCA</i>	<i>MIR4664</i>	<i>POLR3GL</i>	<i>SLFN11</i>	<i>YY1</i>
<i>C10orf88</i>	<i>DNAJB1</i>	<i>HPN-AS1</i>	<i>MIR4672</i>	<i>POMT1</i>	<i>SNRK</i>	<i>ZBP1</i>
<i>C16orf13</i>	<i>DNAJC28</i>	<i>IFNAR2</i>	<i>MIR499A</i>	<i>PRDX5</i>	<i>SOX5</i>	<i>ZNF267</i>
<i>C16orf80</i>	<i>EPHX3</i>	<i>IMP4</i>	<i>MIR6075</i>	<i>PRMT6</i>	<i>SPIRE2</i>	<i>ZNF277</i>
<i>C19orf59</i>	<i>ESAM</i>	<i>INSRR</i>	<i>MIR6743</i>	<i>RANGAP1</i>	<i>STK17A</i>	<i>ZNF555</i>
<i>C1orf50</i>	<i>ESR1</i>	<i>KCTD7</i>	<i>MIR6848</i>	<i>RBM47</i>	<i>STX16</i>	
<i>CACNA1G</i>	<i>FAM115A</i>	<i>KLK14</i>	<i>MKNK1-AS1</i>	<i>RBMXL2</i>	<i>TBC1D1</i>	
<i>CASP7</i>	<i>FAM205A</i>	<i>LARP1</i>	<i>MX2</i>	<i>RNF4</i>	<i>TCF12</i>	
<i>CECR3</i>	<i>FEZ1</i>	<i>LDHD</i>	<i>MYLK3</i>	<i>RP11-31F15.2</i>	<i>TGIF1</i>	

Table S9 – List of genes with hypo-methylated CpGs uniquely identified by the proposed Bayesian model.

<i>ABTB2</i>	<i>CAPN9</i>	<i>FAM218A</i>	<i>LOC100506422</i>	<i>NXT1</i>	<i>RGR</i>	<i>SYT3</i>
<i>AC009473.1</i>	<i>CAST</i>	<i>FAM83D</i>	<i>LOC101928604</i>	<i>OAF</i>	<i>RNF138</i>	<i>TAS1R1</i>
<i>AC010975.1</i>	<i>CBFB</i>	<i>FAM83E</i>	<i>LOC101929217</i>	<i>OCIAD1</i>	<i>RNF139</i>	<i>TBCA</i>
<i>AC097662.2</i>	<i>CCDC167</i>	<i>FCN1</i>	<i>LOC101929528</i>	<i>OR11L1</i>	<i>RNF165</i>	<i>TCTEX1D4</i>
<i>ACOXL</i>	<i>CCDC68</i>	<i>FOXI2</i>	<i>LOC284023</i>	<i>OR13G1</i>	<i>RNLS</i>	<i>TDRD9</i>
<i>ACSM5</i>	<i>CCL16</i>	<i>FOXJ3</i>	<i>LOC440461</i>	<i>OR2AG1</i>	<i>RP11-421P23.1</i>	<i>TEAD3</i>
<i>ADAMTS16</i>	<i>CD27</i>	<i>FZD5</i>	<i>LRCH3</i>	<i>OR56A3</i>	<i>RP11-569G13.2</i>	<i>TECRL</i>
<i>ADAMTSL1</i>	<i>CD47</i>	<i>G2E3</i>	<i>LRLIT1</i>	<i>OR7E91P</i>	<i>RP11-666I19.2</i>	<i>TEKT4</i>
<i>ADD2</i>	<i>CDKN1B</i>	<i>GALC</i>	<i>LRPAP1</i>	<i>OSCAR</i>	<i>RP11-807H7.1</i>	<i>TFAP2A-AS1</i>
<i>AGFG2</i>	<i>CDKN2A</i>	<i>GK</i>	<i>LRRC56</i>	<i>P4HB</i>	<i>RP11-857B24.5</i>	<i>TGFBR1</i>
<i>AGO3</i>	<i>CELSR2</i>	<i>GLIS3</i>	<i>LZTS1-AS1</i>	<i>PALLD</i>	<i>RP11-909B2.1</i>	<i>THOC1</i>
<i>AJUBA</i>	<i>CES3</i>	<i>GML</i>	<i>MALT1</i>	<i>PAOX</i>	<i>RP5-1180C10.2</i>	<i>THY1</i>
<i>AK2</i>	<i>CHIAP2</i>	<i>GNAL</i>	<i>MAN2B2</i>	<i>PCDHAC1</i>	<i>RPL6</i>	<i>TIPRL</i>
<i>AKAP9</i>	<i>CNKSRI</i>	<i>GNL3L</i>	<i>MAP4</i>	<i>PCDHB11</i>	<i>RRAGD</i>	<i>TMEM14C</i>
<i>AKTIP</i>	<i>CNOT3</i>	<i>GPR12</i>	<i>MAPK8IP1</i>	<i>PCDHGB2</i>	<i>RUNX1</i>	<i>TMEM180</i>
<i>ALDH18A1</i>	<i>CNPY3</i>	<i>GPR137C</i>	<i>MARCKS</i>	<i>PCSK2</i>	<i>S100A11</i>	<i>TMEM41B</i>
<i>ALG5</i>	<i>COCH</i>	<i>GPR156</i>	<i>MC5R</i>	<i>PELO</i>	<i>SAG</i>	<i>TRAF3IP1</i>
<i>ALOX5AP</i>	<i>COL18A1-AS2</i>	<i>GPR6</i>	<i>MCL1</i>	<i>PGS1</i>	<i>SAMD12-AS1</i>	<i>TRAM2</i>
<i>ALPP</i>	<i>CPEB1</i>	<i>GPR62</i>	<i>MEGF11</i>	<i>PHOSPHO1</i>	<i>SAMD9</i>	<i>TRIM8</i>
<i>AMH</i>	<i>CSAD</i>	<i>GRIA2</i>	<i>MEP1A</i>	<i>PHYHIP</i>	<i>SCARB2</i>	<i>TRPC4</i>
<i>ANAPC1</i>	<i>CSRNP1</i>	<i>GS1-259H13.2</i>	<i>MGC34796</i>	<i>PIGB</i>	<i>SCYL3</i>	<i>TRPM5</i>
<i>ANO2</i>	<i>CSTF2</i>	<i>GXYLT2</i>	<i>MIR10B</i>	<i>PIH1D3</i>	<i>SEMA5B</i>	<i>TUBGCP5</i>
<i>AP1AR</i>	<i>CUL4B</i>	<i>HDAC10</i>	<i>MIR340</i>	<i>PKD1L2</i>	<i>SERPINA7</i>	<i>TUSC1</i>
<i>AP5S1</i>	<i>CYP4F2</i>	<i>HIST3H2A</i>	<i>MIR365A</i>	<i>PLA2G4C</i>	<i>SGSM1</i>	<i>UBE2B</i>
<i>ARHGAP20</i>	<i>DAW1</i>	<i>HMGA1</i>	<i>MIR3688-2</i>	<i>PLD4</i>	<i>SH2B1</i>	<i>UBP1</i>
<i>ARHGDI1B</i>	<i>DBNDD1</i>	<i>HSPA4</i>	<i>MIR3935</i>	<i>PLEKHA1</i>	<i>SHC2</i>	<i>ULK4</i>
<i>ATP5H</i>	<i>DBP</i>	<i>HSPA8</i>	<i>MIR4533</i>	<i>POU3F3</i>	<i>SKOR1</i>	<i>UNKL</i>
<i>ATP7B</i>	<i>DBX1</i>	<i>HSPB3</i>	<i>MIR4649</i>	<i>PPCDC</i>	<i>SLC13A1</i>	<i>UQCRCFS1</i>
<i>AZU1</i>	<i>DCAF7</i>	<i>HTR1B</i>	<i>MIR4694</i>	<i>PPM1A</i>	<i>SLC25A20</i>	<i>UQCRL</i>
<i>BAG1</i>	<i>DDX49</i>	<i>INTS5</i>	<i>MIR5091</i>	<i>PRICKLE4</i>	<i>SLC25A51</i>	<i>USP46</i>
<i>BAHD1</i>	<i>DIAPH1</i>	<i>IRX5</i>	<i>MIR548AS</i>	<i>PRKG1</i>	<i>SLC27A6</i>	<i>VIPR1</i>
<i>BCL11A</i>	<i>DLST</i>	<i>ISX</i>	<i>MIR548I1</i>	<i>PRMT1</i>	<i>SLC28A3</i>	<i>VSTM2B</i>
<i>BCL6</i>	<i>DLX2</i>	<i>ITGA8</i>	<i>MIR6795</i>	<i>PRR14L</i>	<i>SLC29A1</i>	<i>WBP11P1</i>
<i>BEAN1</i>	<i>DNAJC1</i>	<i>ITPRIPL2</i>	<i>MIR6867</i>	<i>PRSS22</i>	<i>SLC37A3</i>	<i>WDYHV1</i>
<i>BID</i>	<i>DNMT3B</i>	<i>IWS1</i>	<i>MKLN1</i>	<i>PRSS38</i>	<i>SLC46A1</i>	<i>WNT16</i>
<i>BMP4</i>	<i>DPEP2</i>	<i>KCNIP3</i>	<i>MKRN1</i>	<i>PSIP1</i>	<i>SLC4A8</i>	<i>YWHAB</i>
<i>BOC</i>	<i>DPH2</i>	<i>KIAA1024L</i>	<i>MRPL12</i>	<i>PSMA3</i>	<i>SLC6A12</i>	<i>ZFP36</i>
<i>BPIFC</i>	<i>DPP10</i>	<i>KLHDC10</i>	<i>MSI1</i>	<i>PSMC3</i>	<i>SMCO4</i>	<i>ZNF181</i>
<i>BRSK1</i>	<i>DPY19L2</i>	<i>KLK5</i>	<i>MTHFD2</i>	<i>PTAFR</i>	<i>SMKR1</i>	<i>ZNF184</i>
<i>BST2</i>	<i>DSP</i>	<i>LILRB2</i>	<i>MVD</i>	<i>PTBP3</i>	<i>SNX22</i>	<i>ZNF236</i>
<i>C10orf126</i>	<i>DUOXA1</i>	<i>LINC00442</i>	<i>NASP</i>	<i>PTPMT1</i>	<i>SPATA41</i>	<i>ZNF488</i>
<i>C10orf95</i>	<i>EDNRB</i>	<i>LINC00535</i>	<i>NBPF23P</i>	<i>PVRL1</i>	<i>SPDYA</i>	<i>ZNF496</i>
<i>C12orf45</i>	<i>EFR3B</i>	<i>LINC00644</i>	<i>NCF4</i>	<i>RAB27A</i>	<i>SRSF12</i>	<i>ZNF550</i>
<i>C1orf234</i>	<i>EGFLAM-AS2</i>	<i>LINC00700</i>	<i>NCR1</i>	<i>RAD54L2</i>	<i>SRSF4</i>	<i>ZNF606</i>
<i>C20orf27</i>	<i>EGFR</i>	<i>LINC00707</i>	<i>NDP</i>	<i>RAP1GAP2</i>	<i>SSNA1</i>	<i>ZNF718</i>
<i>C2orf50</i>	<i>EIF5</i>	<i>LINC00896</i>	<i>NDUFA11</i>	<i>RAP2B</i>	<i>ST6GALNAC6</i>	<i>ZNF768</i>
<i>C6orf132</i>	<i>EPB41L4A</i>	<i>LINC01003</i>	<i>NECAB2</i>	<i>RASA1</i>	<i>STAP2</i>	<i>ZNF99</i>
<i>C8B</i>	<i>ETFA</i>	<i>LINC01007</i>	<i>NFATC3</i>	<i>RASA4CP</i>	<i>STK25</i>	
<i>C8orf48</i>	<i>FAM118A</i>	<i>LINC01056</i>	<i>NOX1</i>	<i>RASGRP1</i>	<i>STMN2</i>	
<i>C8orf86</i>	<i>FAM126A</i>	<i>LINC01116</i>	<i>NR5A2</i>	<i>RAX2</i>	<i>STRA6</i>	
<i>CAMK1</i>	<i>FAM184B</i>	<i>LIPG</i>	<i>NRF1</i>	<i>RBM20</i>	<i>SUDS3</i>	
<i>CANT1</i>	<i>FAM213A</i>	<i>LMNB2</i>	<i>NRP1</i>	<i>RGCC</i>	<i>SYCE1L</i>	
<i>CAPN10-AS1</i>	<i>FAM215A</i>	<i>LOC100268168</i>	<i>NTN4</i>	<i>RGPD3</i>	<i>SYNJ1</i>	

Figures

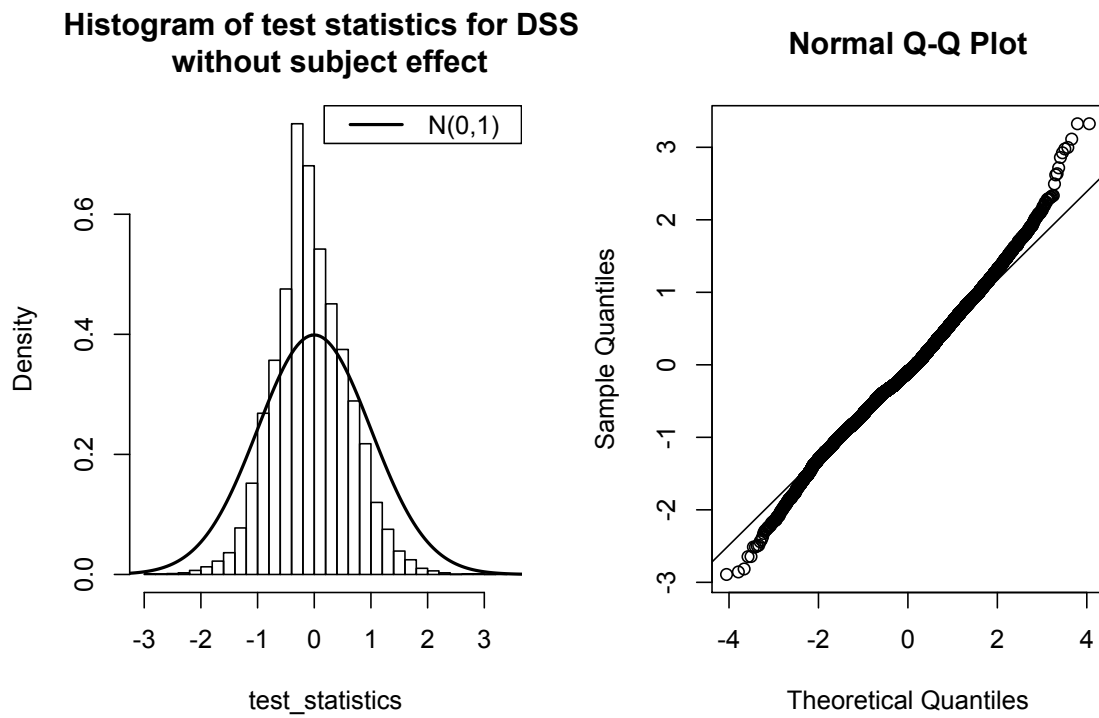


Figure S1 – Left panel: the histogram of the empirical Bayes Wald test statistics of the DSS procedure from one simulated data without subject effect compared with the standard normal curve. Right panel: the normal QQ plot of the empirical Bayesian Wald test statistics of one simulated data without subject effect.

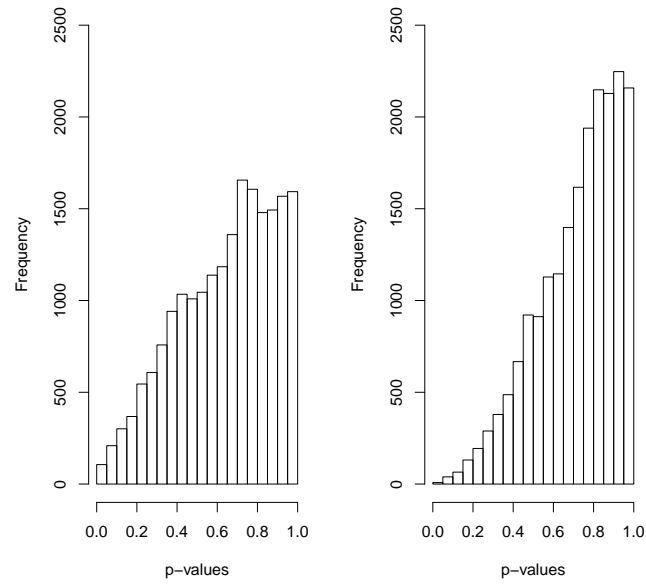


Figure S2 – Histograms of p -values for the empirical Bayes Wald test on one dataset simulated under the null hypothesis (i.e. all 20,000 CpG loci are simulated as equal-methylated). Left panel is for the first set of simulation studies (without subject effect) and right panel is for the second set of simulation studies (with subject effect).

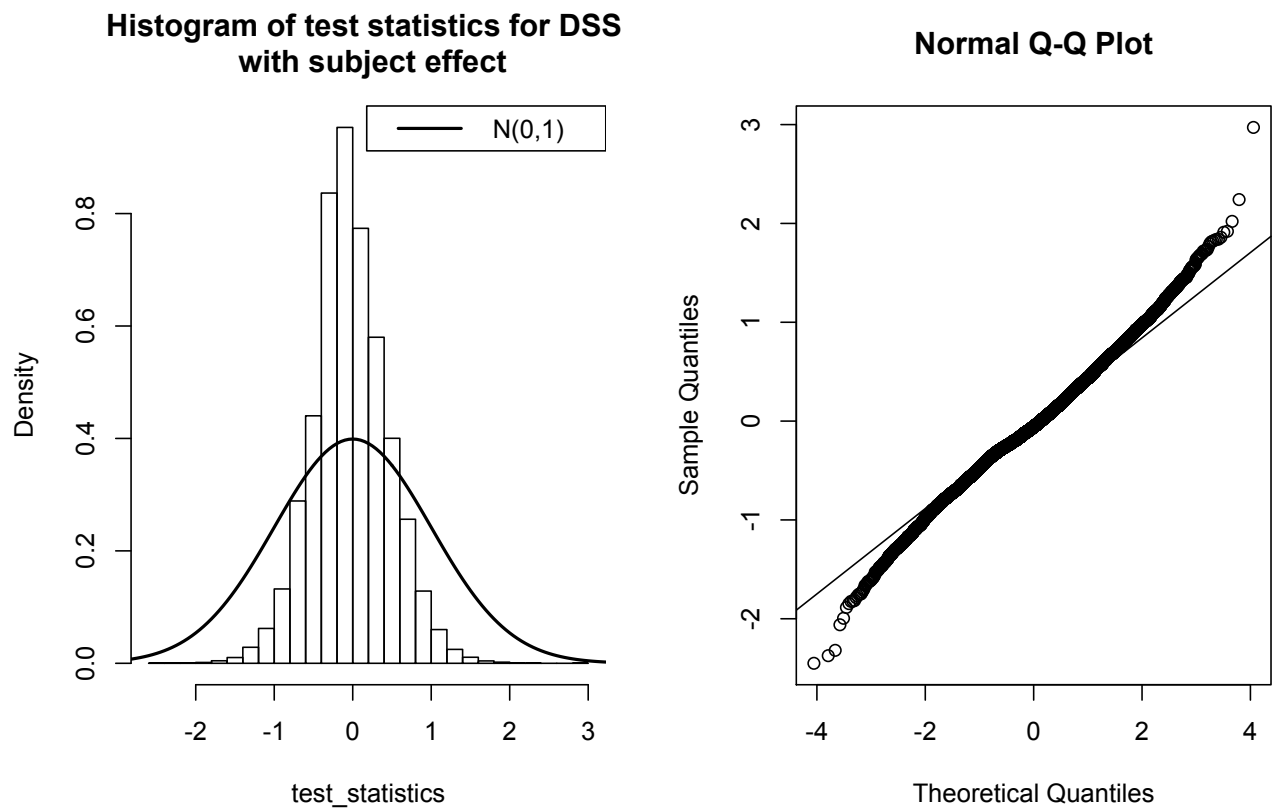


Figure S3 – Similar to Figure S1, except that this is for the simulation studies with subject effect.

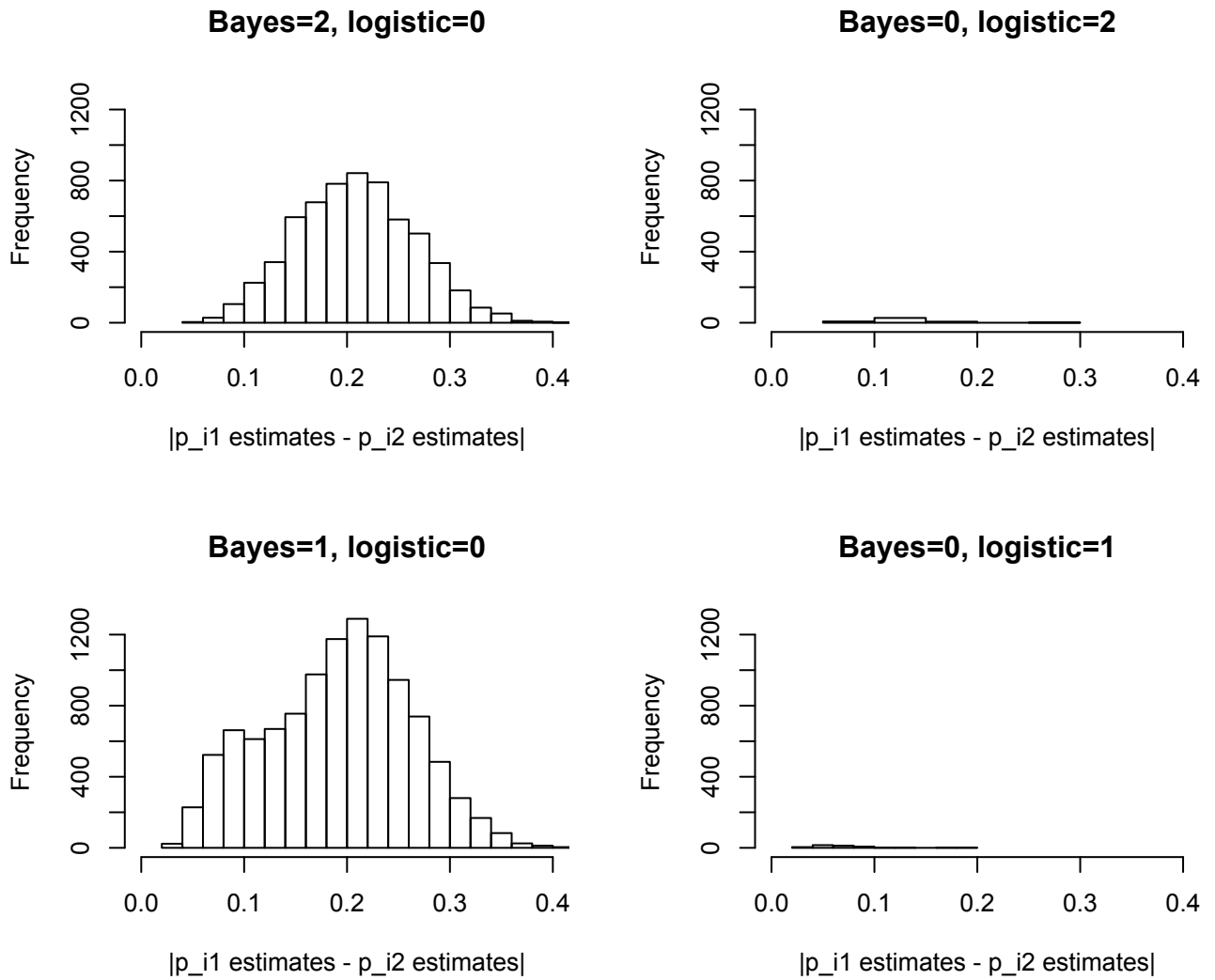


Figure S4 – For the loci uniquely identified to be hypo-methylated loci (group 1) or hyper-methylated loci (group 2) by either the proposed Bayesian method or logistic regression when applied to the real data, four panels are histograms of absolute difference for methylation proportion estimates between the case and control samples. These four panels correspond to the counts in Table 4 when FDR level for logistic regression is controlled at 0.01 level.

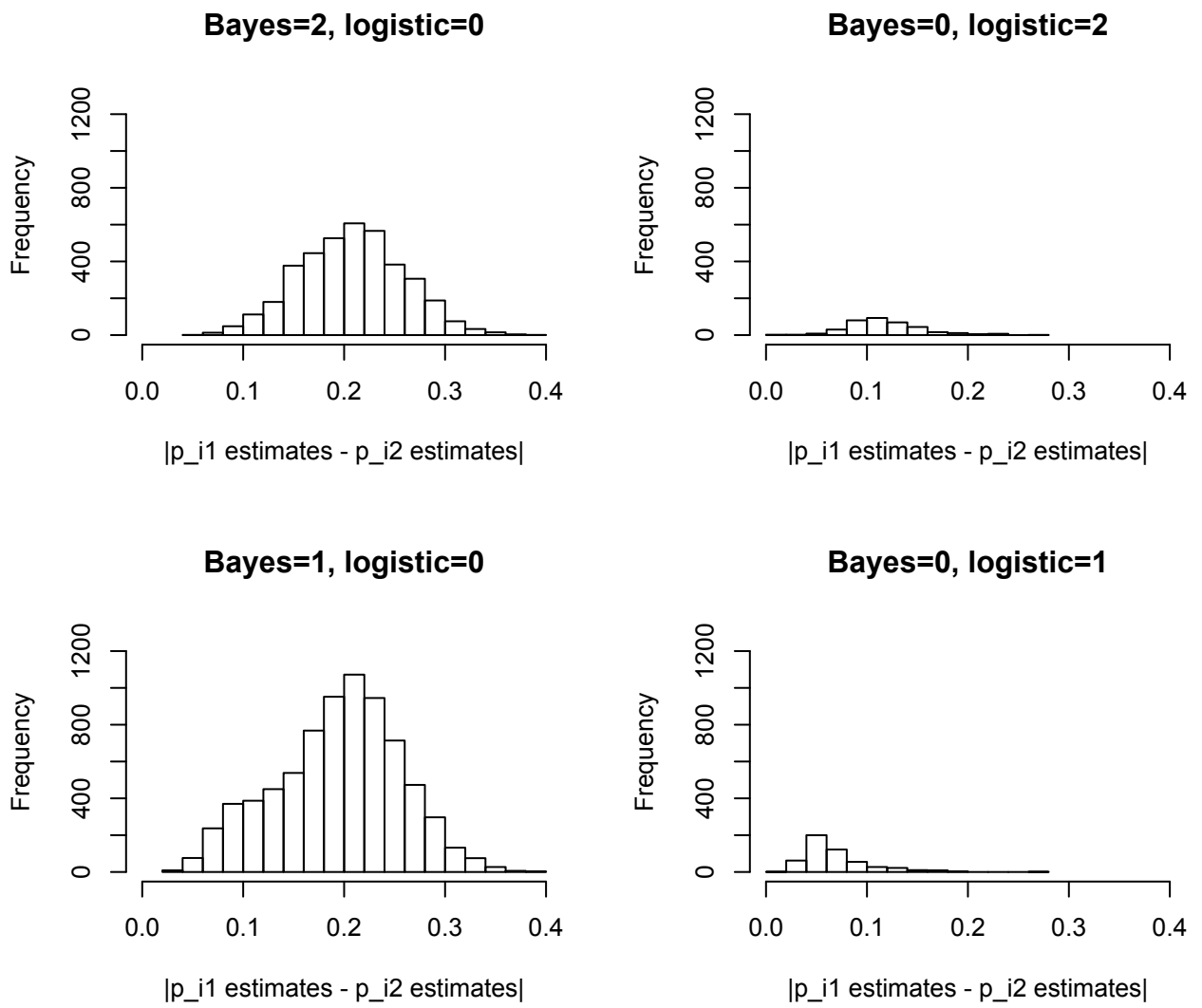


Figure S5 – Similar to Figure S4, except that FDR level for logistic regression is controlled at 0.02.

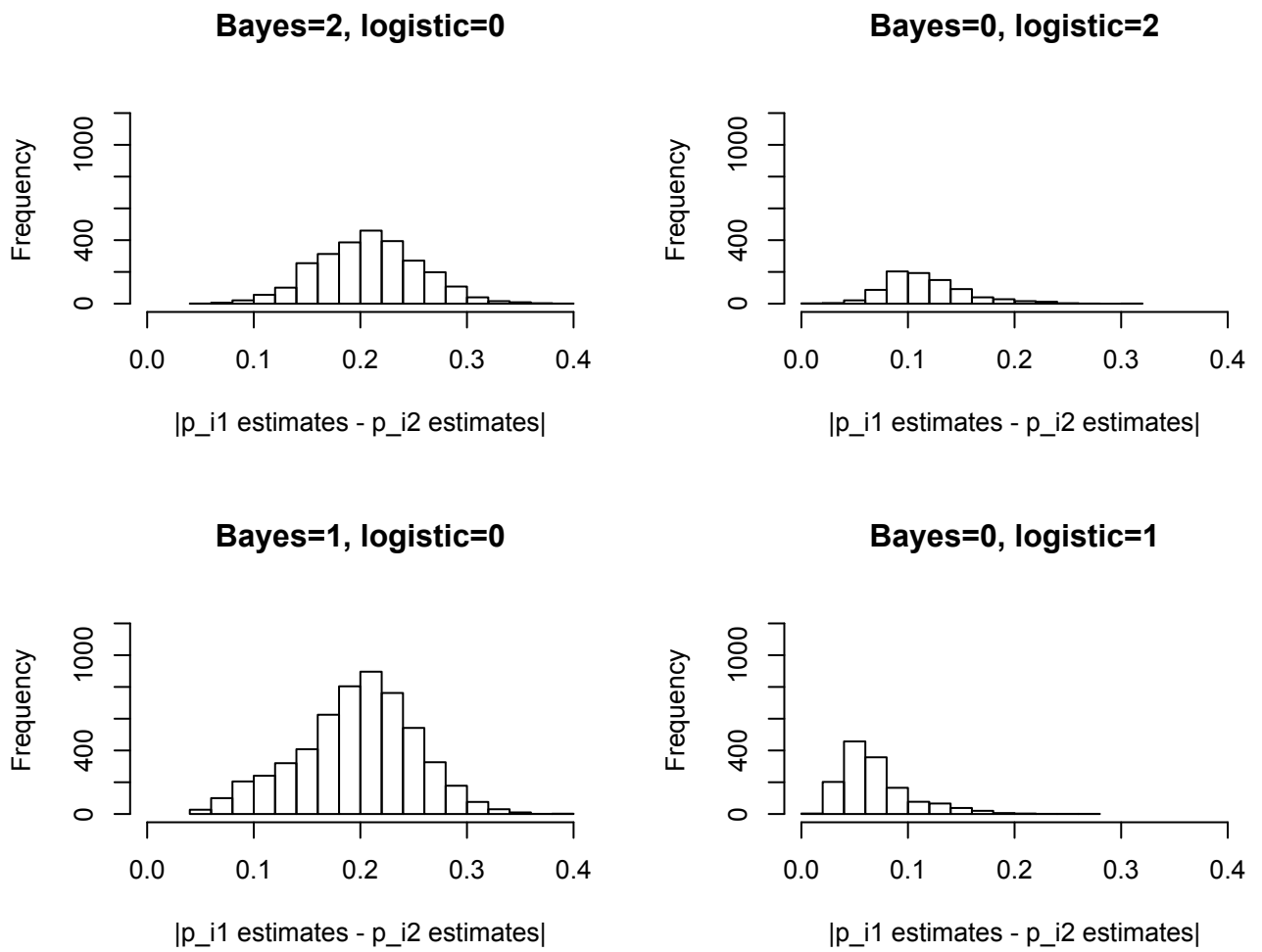
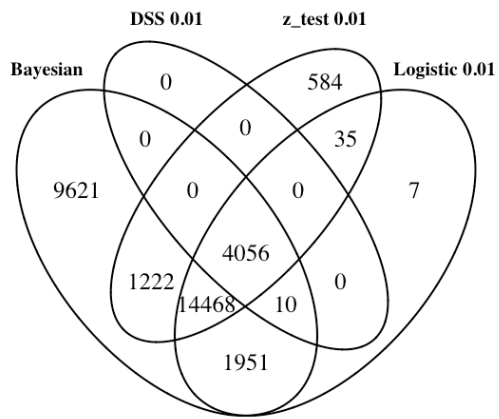
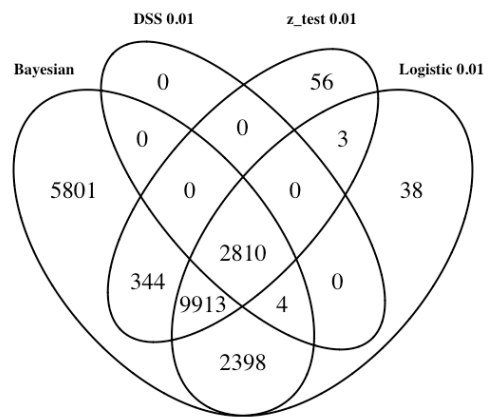


Figure S6 – Similar to Figure S4, except that FDR level for logistic regression is controlled at 0.03.

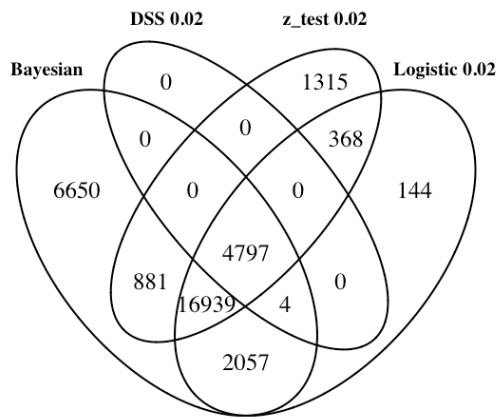


(a) Hypo-methylated

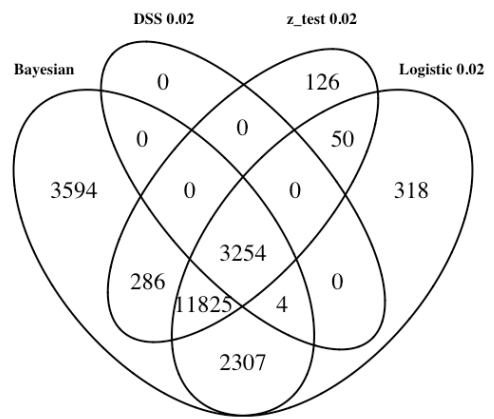


(b) Hyper-methylated

Figure S7 – Venn diagram of detected hypo- and hyper-methylated loci in real data analysis. FDR for logistic regression, DSS and z test is controlled at 0.01 level.

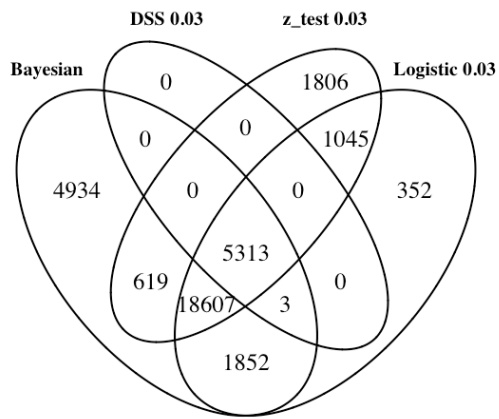


(a) Hypo-methylated

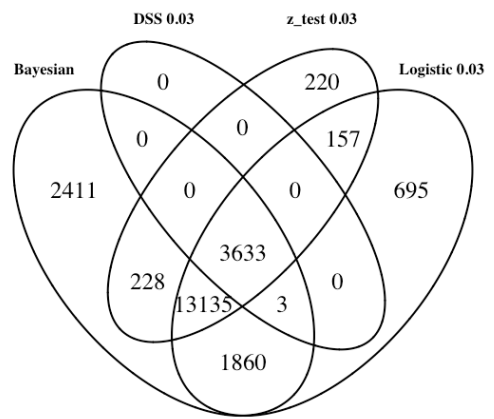


(b) Hyper-methylated

Figure S8 – Venn diagram of detected hypo- and hyper-methylated loci in real data analysis. FDR for logistic regression, DSS and z test is controlled at 0.02 level.

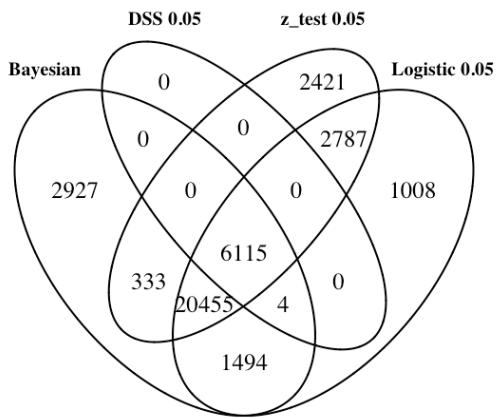


(a) Hypo-methylated

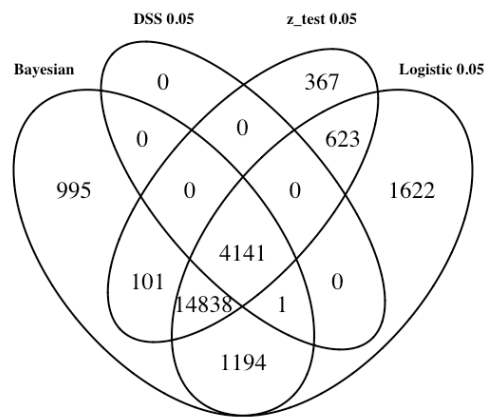


(b) Hyper-methylated

Figure S9 – Venn diagram of detected hypo- and hyper-methylated loci in real data analysis. FDR for logistic regression, DSS and z test is controlled at 0.03 level.



(a) Hypo-methylated



(b) Hyper-methylated

Figure S10 – Venn diagram of detected hypo- and hyper-methylated loci in real data analysis. FDR for logistic regression, DSS and z test is controlled at 0.05 level.

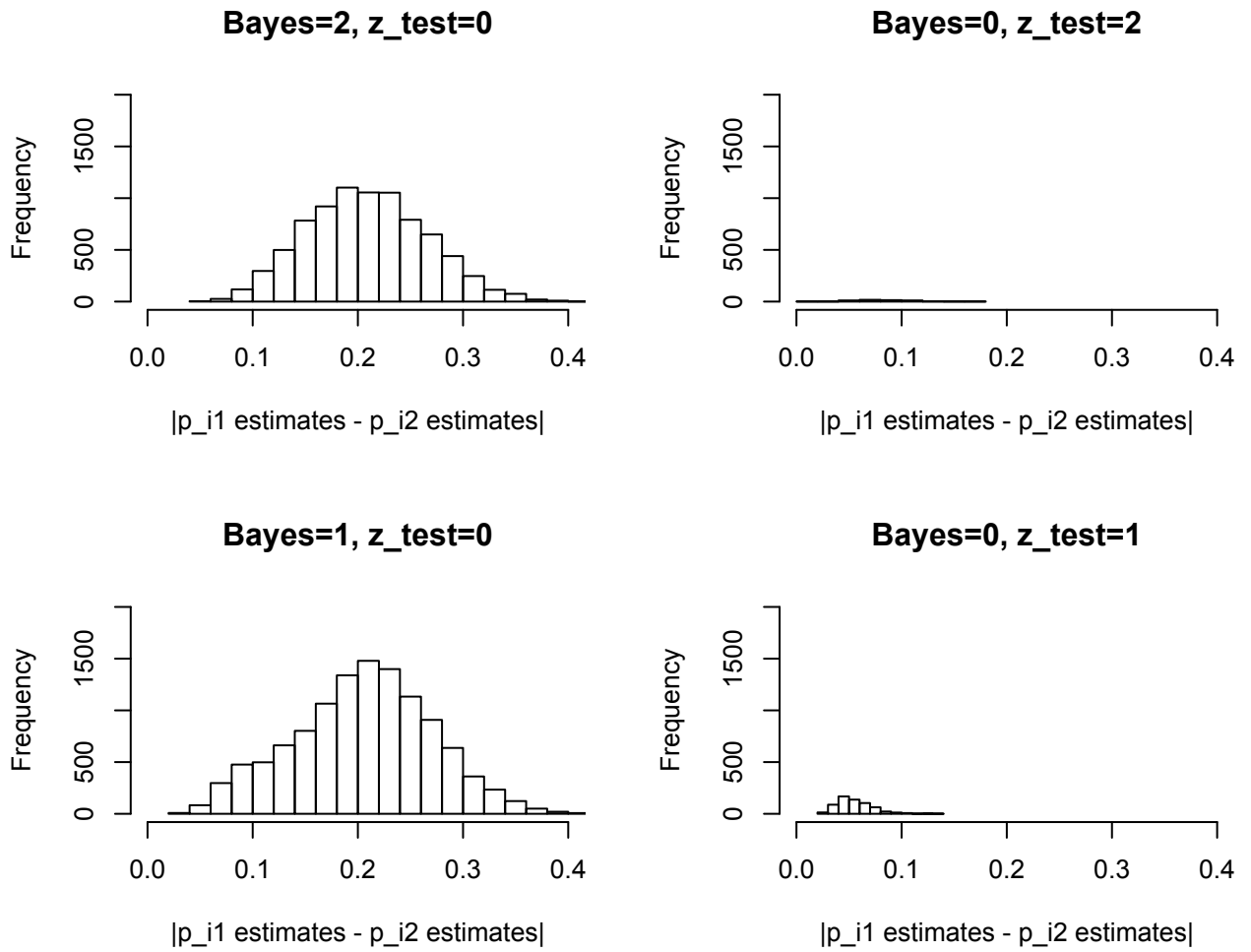


Figure S11 – For the loci uniquely identified to be hypo-methylated loci (group 1) or hyper-methylated loci (group 2) by either the proposed Bayesian method or z test when applied to the real data, four panels are histograms of absolute difference for methylation proportion estimates between the case and control samples. FDR level for z test is controlled at 0.01.

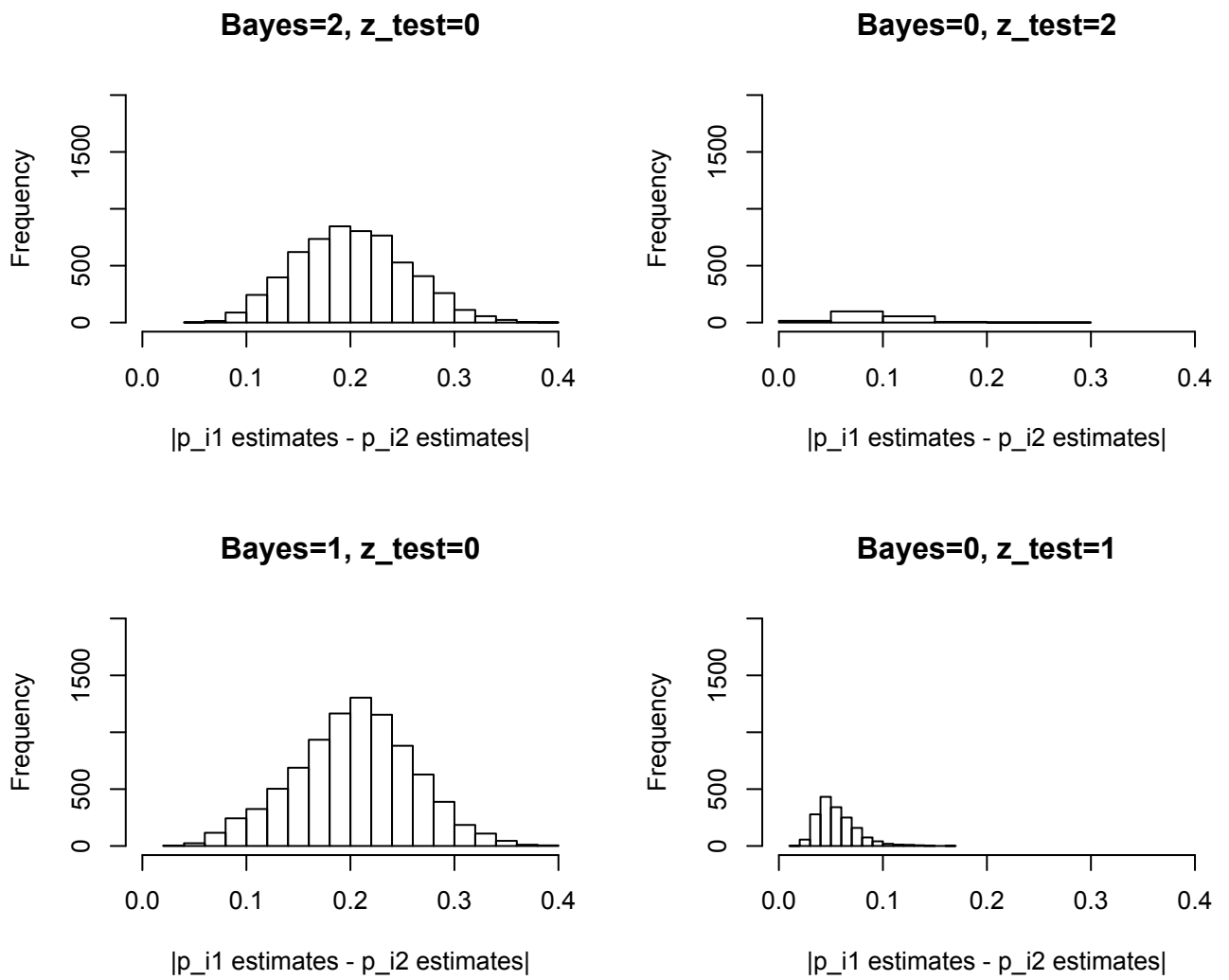


Figure S12 – Similar to Figure S11, except that FDR level for z test is controlled at 0.02.

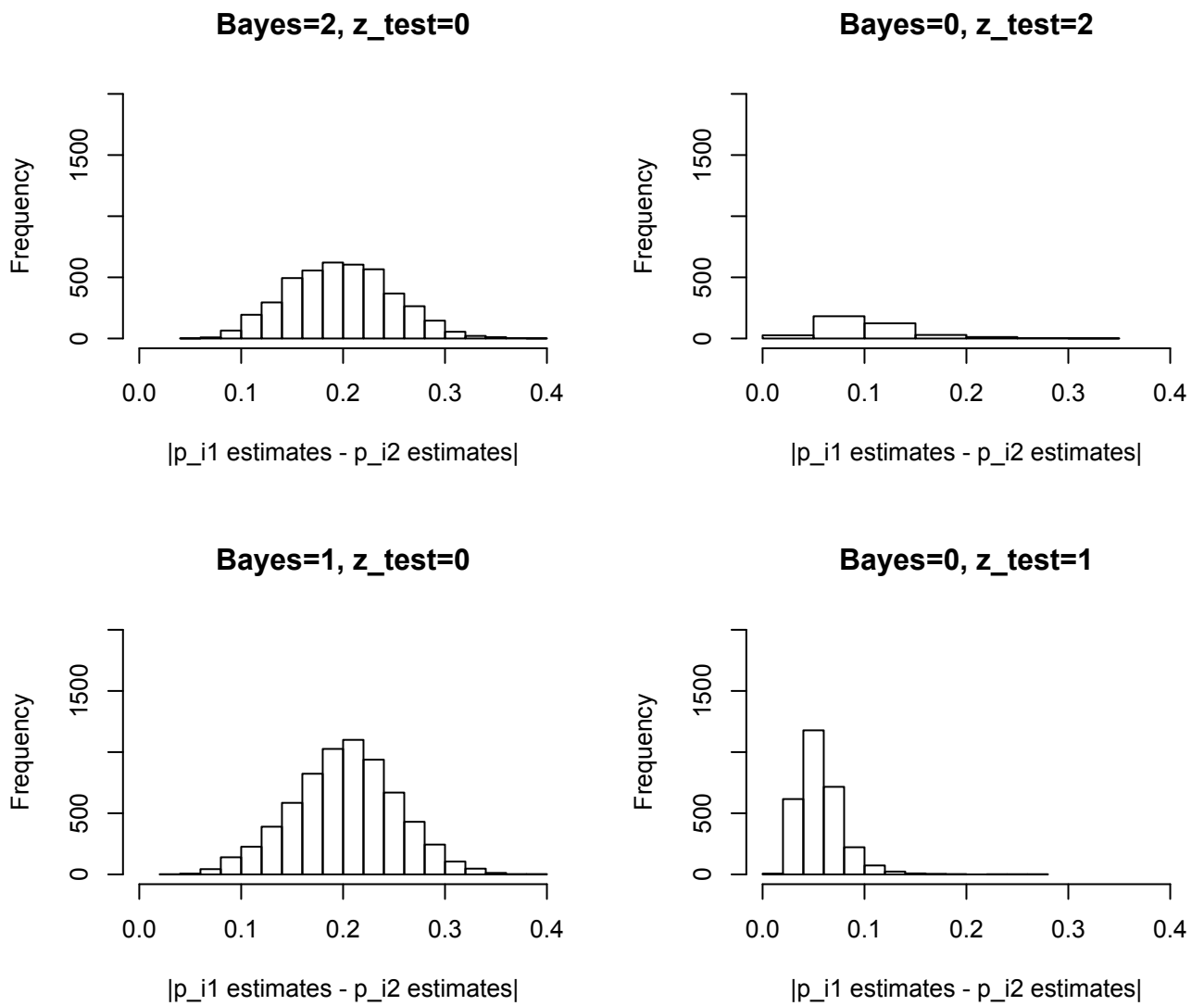


Figure S13 – Similar to Figure S11, except that FDR level for z test is controlled at 0.03.

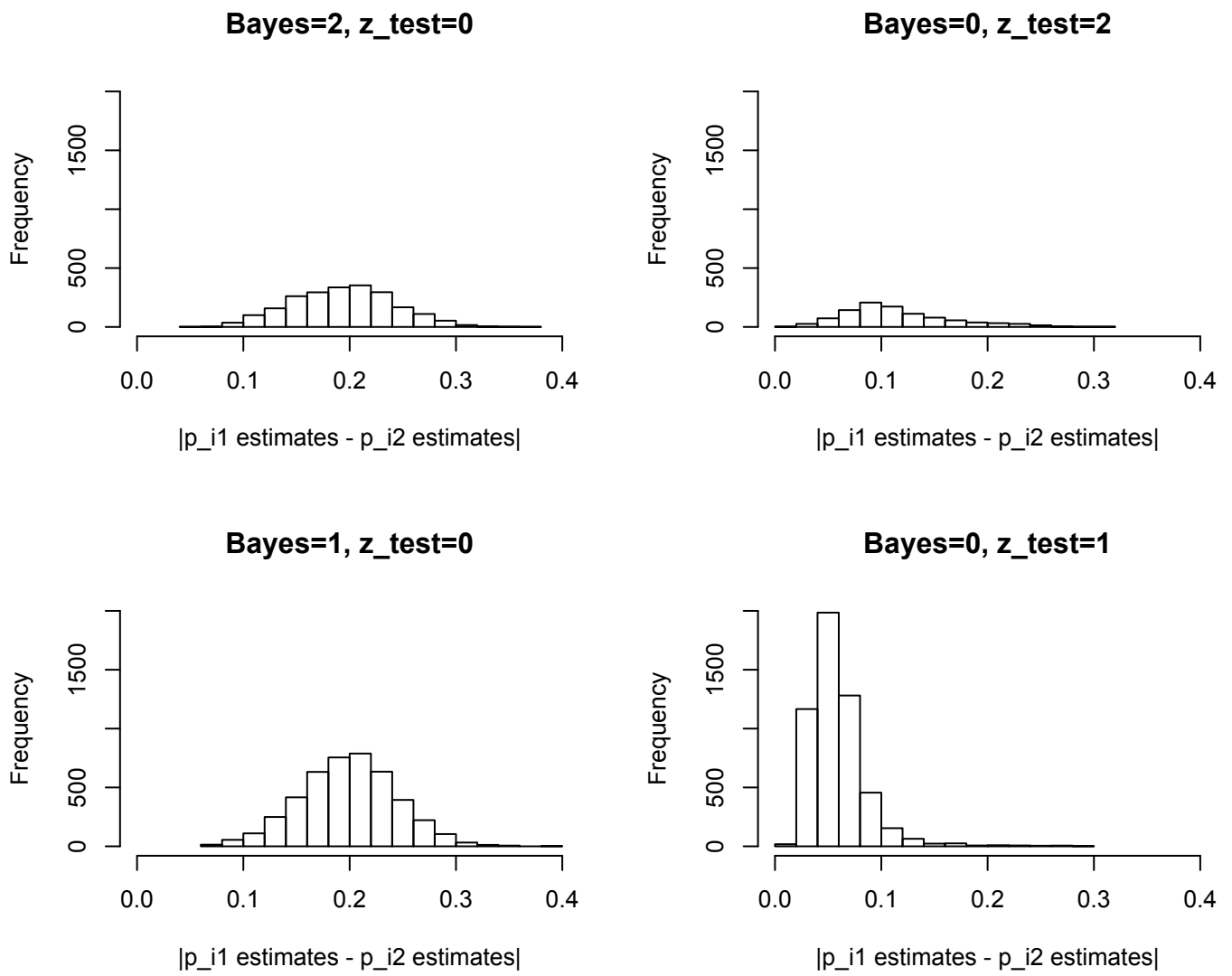


Figure S14 – Similar to Figure S11, except that FDR level for z test is controlled at 0.05.