Methods

Modelization of the methylation level of each region

Among 11 region categories, included in the 450k BeadChip annotation, the promoter, 5' UTR, 1st exon, gene body and 3' UTR are always associated to a gene whereas the CpG island category and its surrounding shore and shelve regions may not be, depending on their distance to the nearest genes. As each region contains at least one CpG site, treating the region as a unit in the differential methylation analysis might help identify regions with consistently coordinate methylation changes. From a statistical point of view, region-based differential methylation analysis reduces the burden of multiple comparisons and increases the power to catch differentially methylated regions associated with SMA (1). For each specific region the methylation levels (M-values) of the CpG sites associated with the same gene/CpG island were averaged.

Results

Analysis at the CpG sites level

Among these genes were *MAFK* and *FBXL7,* for which both groups' significant CpG sites belonged to the same CpG island. Furthermore, for *COLEC11*, *HLA-G*, *GALNTL4* and *TRIM26* more than one CpG site with a consistent fold change was found to be significant. *NCRNA00171*, *KIAA1217* and *DIP2C* had more complex methylation patterns, as at least two sites, associated with each gene, had opposite fold changes in one group. In the case of *KCNQ1*, there was one CpG site located in the gene body with a negative fold change for both analyzed groups. Six CpG sites belonging to a common CpG island, and located within 1500 bp upstream of the

transcription start site (TSS), had a positive fold change in second analyzed group. However, these 6 CpG sites were located close to *KCNQ1DN* (KCNQ1 Downstream Neighbor), which contains a long non-coding RNA gene subject to imprinting.

Analysis at the region level

For each of the 11 annotated regions, the link between the methylation profile and SMA was identified by fitting a linear model that explains the average methylation level for each gene/CpG island. In second analyzed group, the average methylation level of CpG sites within 1500 bp upstream of the *HLA-G* gene TSS was significantly higher in controls than in patients, which was consistent with the methylation level of the significant CpG site found in the body of *HLA-G* (Supplementary table 1). The south shore of a CpG island associated with the *AP2A2* gene showed a significant negative fold change, which was in line with the methylation level of the significant CpG site located in *AP2A2*'s 3' UTR. A CpG island associated with *MTUS2* showed a strong and consistent pattern in both groups, the average methylation levels of the island itself as well as its North shelf were significantly higher in controls than in patients. There was another gene with a complex methylation pattern: in first analyzed group, the average methylation level of an island associated with *CACNA1C* displayed a significant positive fold change whereas a unique CpG site located in its 3'UTR displayed an opposite fold change.

Discussion

One CpG site is located inside the region of chromosome 13 open reading frame 16 (C13orf16), with uncharacterized protein C13orf16 encoded (http://genome.ucsc.edu/ GRCh37/hg19). We could hypothesize that these intragenic CpG sites are associated with transcriptional start sites (TSS) of noncoding RNAs (ncRNAs), and thus regulating the expression of several genes (2). To note, three

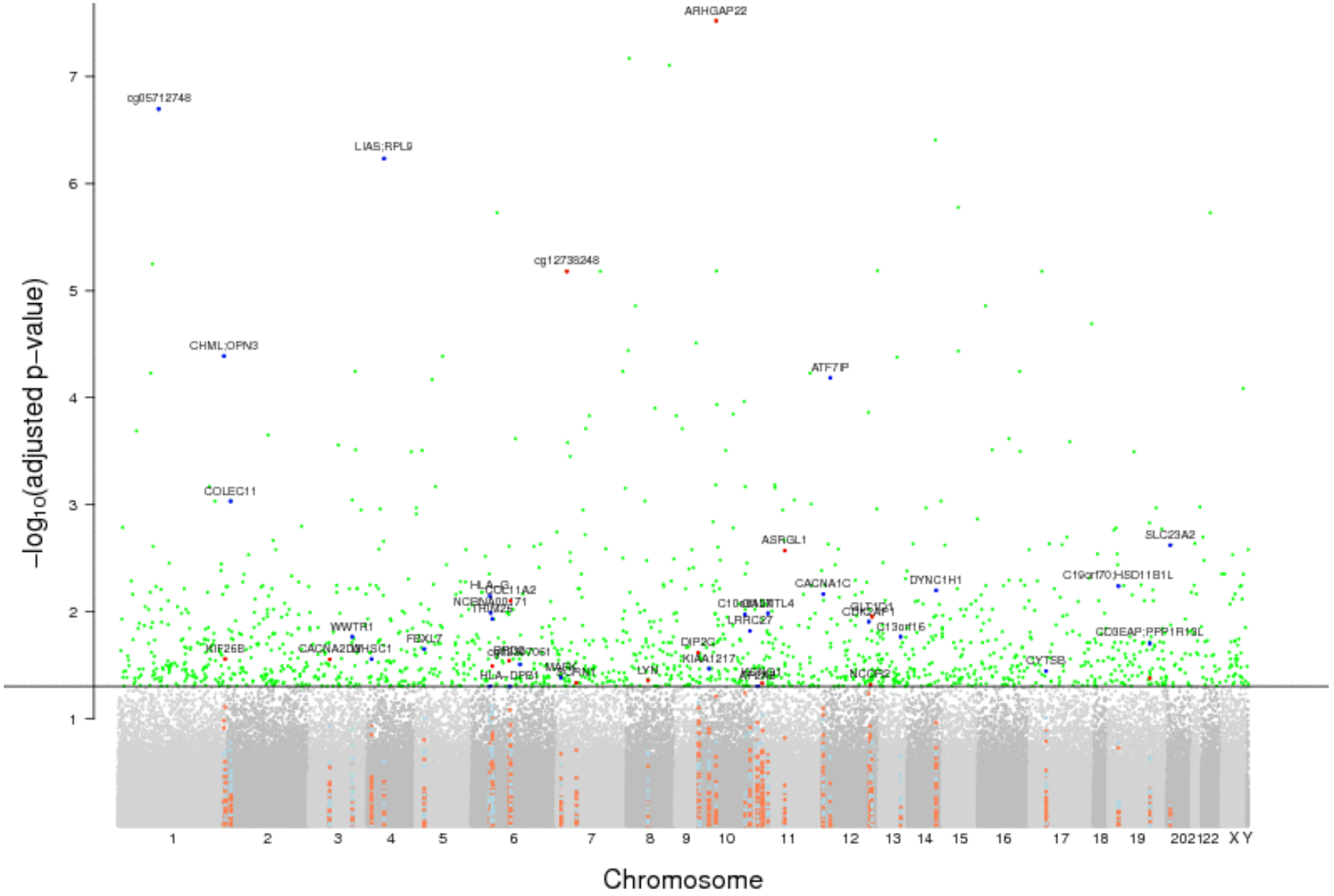
from six described above CpG sites are located near TSS in promoter regions of gene that supports their direct influence on the process of transcription intensity. Three other CpG sites are localized in intergenic areas (Table 2). They can also be associated with additional TSS for regulatory ncRNAs, and thus may affect transcriptional elongation or alternative splicing of genes (2).

Differently methylated CpG sites were also detected in the *DYNC1H1* and *KIF26B* genes, which encode heavy chain 1 of cytoplasmic dynein 1 and kinesin family member 26B, respectively. Strong differences in methylation levels were found for a CpG island connected with *MTUS2* gene, which may participate in the regulation of microtubule dynamics at their growing distal tip (3). The products of *AP2A2*, *CACNA1C* and *MAFK* genes participate in axon guidance processes and in nervous system development in general (string-db.org, GO enrichment analysis, medium confidence score 0,400). It is of interest to note the regulatory role of PPP1R13L protein in apoptosis via its interaction with NF-kappa-B and p53/TP53 proteins (www.genecards.org).

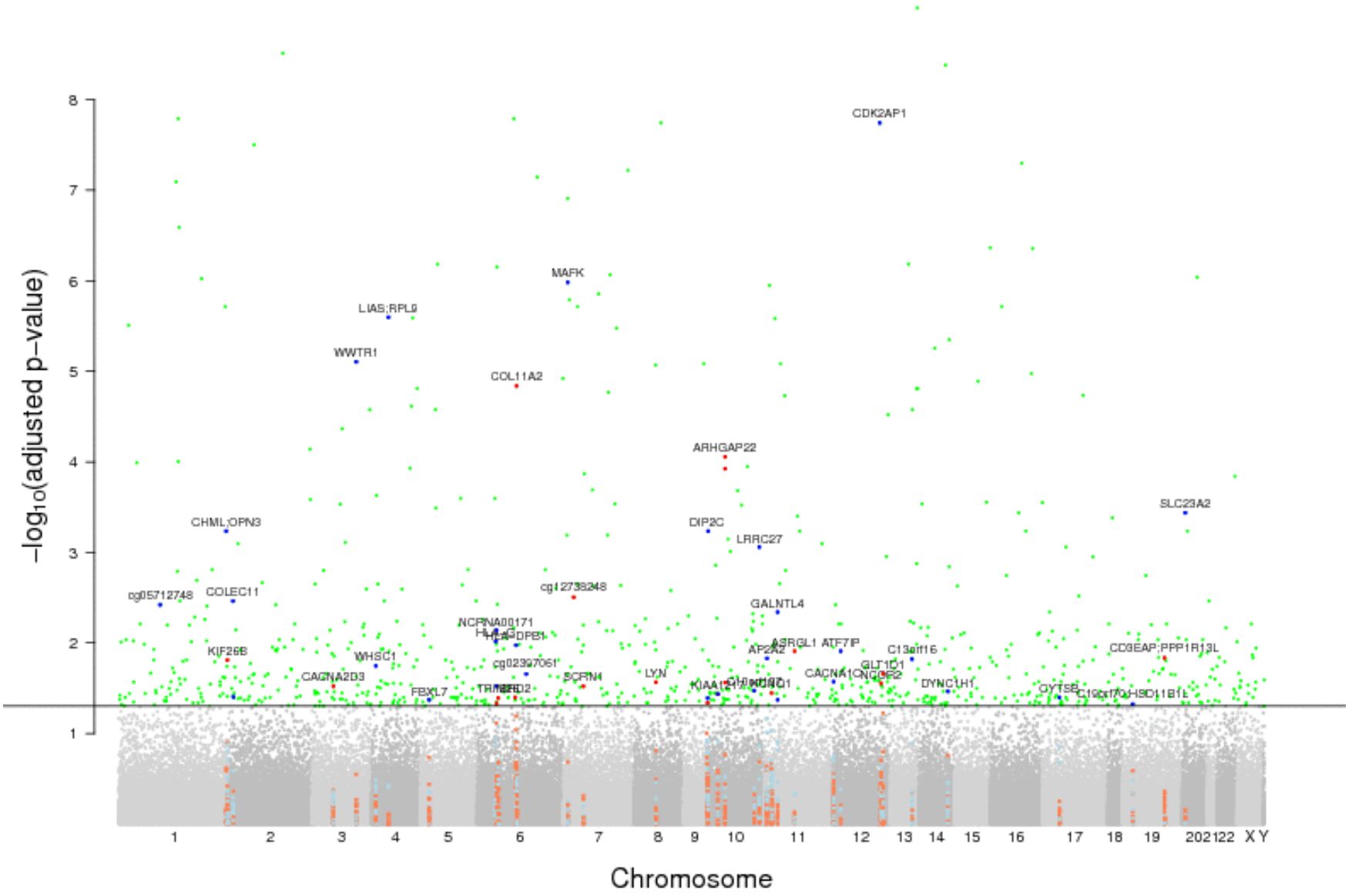
References

1. Wang D, Yan L, Hu Q, et al: IMA: An R Package for High-Throughput Analysis of Illumina's 450K Infinium Methylation Data. *Bioinformatics* [Internet]. 2012.
2. Deaton AM, Bird A: CpG islands and the regulation of transcription. *Genes Dev* 2011; **25(10):**1010–22.
3. Jiang K, Wang J, Liu J et al: TIP150 interacts with and targets MCAK at the microtubule plus ends. *EMBO Rep* 2009; **10(8):** 857–865.

Supplementary Figure 1A



Supplementary Figure 1B



Supplementary figure 1A, 1B. Manhattan plots of CpG sites/genes for first (A) and second (B) groups. Manhattan plots show genomic coordinates vs the negative logarithm of adjusted p-value for each CpG site. The horizontal black line indicates the significance threshold (p-value = 0.05), significant points above this line are green. Common significant methylated CpG sites/genes are highlighted. The sites more methylated in controls than in patients are displayed in blue and in red if less methylated in controls than in patients. Additional non-significant CpG sites associated with significant methylated (highlighted) genes are displayed in light blue if more methylated in controls than in patients and in orange if less methylated in controls than in patients.

Supplementary table 1. Information about significantly differentially methylated regions found for common genes in (I) first (top) and (II) second (bottom) analyzed groups.

Gene	Annotated region	Genomic location	Average β-value (controls)	Average β-value (patients)	Adjusted p-value
MAFK	5'UTR	chr7:1571105	0.0275	0.0261	4.15E-02
WWTR1	Body	chr3:149374763	0.833	0.697	1.72E-02
LRRC27	TSS200	chr10:134145580	0.00958	0.00446	1.51E-02
COLEC11	TSS200	chr2:3642629	0.635	0.503	9.31E-04
GALNTL4	3'UTR	chr11:11292635	0.755	0.496	1.04E-02
ZNRD1-AS1	Body	chr6:29990158	0.945	0.927	1.03E-02
HLA-G	Body	chr6:29796770	0.641	0.505	7.14E-03
HLA-G	Body	chr6:29796593	0.642	0.581	4.96E-02
HLA-DPB1	Body	chr6:33048592	0.702	0.658	5.00E-02
ATF7IP	TSS200	chr12:14518548	0.401	0.111	6.52E-05
AP2A2	Body	chr11:1001560	0.685	0.49	4.96E-02
WHSC1	3'UTR	chr4:1949985	0.967	0.957	2.76E-02
CACNA1C	3'UTR	chr12:2800909	0.00931	0.00504	6.85E-03

CACNA1C	Island	chr12:2788614-2788879	0.987	0.989	3.51E-02
C10orf137	1st Exon	chr10:127408397	0.0703	0.046	1.07E-02
DYNC1H1	Body	chr14:102482777	0.989	0.98	6.31E-03
KIAA1217	TSS1500	chr10:23982350	0.0905	0.0937	3.39E-02
FBXL7	TSS200	chr5:15500270	0.0243	0.0178	2.24E-02
C19orf70/HSD11B1L	1st Exon	chr19:5681171	0.0342	0.0232	5.77E-03
COL11A2	Body	chr6:33131560	0.97	0.978	7.91E-03
ASRGL1	Body	chr11:62138845	0.655	0.727	2.69E-03
CD3EAP/PPP1R13L	5'UTR	chr19:45909395	0.0147	0.012	4.20E-02
KIF26B	Body	chr1:245320016	0.0117	0.0135	2.76E-02
GLT1D1	TSS200	chr12:129337910	0.0663	0.0673	1.11E-02
LYN	TSS200	chr8:56792216	0.0426	0.0525	4.40E-02
NCOR2	5'UTR	chr12:125017134	0.792	0.823	4.85E-02
CACNA2D3	TSS1500	chr3:54155978	0.14	0.147	2.78E-02
SCRN1	TSS1500	chr7:30029938	0.895	0.919	4.63E-02
KCNQ1	TSS1500	chr11:2890419	0.292	0.221	1.95E-02
KCNQ1	TSS1500	chr11:2890647	0.159	0.0963	7.83E-03
KCNQ1	TSS1500	chr11:2890649	0.124	0.0854	3.30E-02
KCNQ1	TSS1500	chr11:2890394	0.137	0.0951	8.83E-03
KCNQ1	TSS1500	chr11:2890668	0.29	0.181	1.18E-02
KCNQ1	TSS1500	chr11:2890670	0.379	0.24	4.18E-02
KCNQ1	Body	chr11:2595605	0.98	0.984	4.68E-02
BRD2	5'UTR	chr6:32939295	0.0777	0.106	2.86E-02
TRIM26	TSS200	chr6:30181175	0.00445	0.00586	1.17E-02
TRIM26	Body	chr6:30166498	0.989	0.993	3.22E-02
DIP2C	Body	chr10:372859	0.916	0.932	2.41E-02
MTUS2	Island	chr13:29913886-29914301	0.849	0.837	4.12E-02
MTUS2	NSHELF	chr13:29913886-29914301	0.734	0.720	3.65E-02

Gene	Annotated region	Genomic location	Average β-value (controls)	Average β-value (patients)	Adjusted p-value
MAFK	5'UTR	chr7:1570407	0.00386	0.000778	1.04E-06
WWTR1	TSS200	chr3:149421196	0.827	0.684	7.88E-06
LRRC27	3'UTR	chr10:134165390	0.826	0.683	8.79E-04
COLEC11	5'UTR	chr2:3646196	0.944	0.87	3.44E-03
COLEC11	5'UTR	chr2:3646388	0.632	0.546	3.96E-02
GALNTL4	Body	chr11:11642360	0.154	0.103	4.27E-02
GALNTL4	TSS200	chr11:11643754	0.0665	0.036	4.58E-03
ZNRD1-AS1	Body	chr6:29969794	0.761	0.634	3.01E-02
ZNRD1-AS1	Body	chr6:30002859	0.887	0.627	7.28E-03
ZNRD1-AS1	Body	chr6:29974991	0.0847	0.102	4.72E-02
HLA-G	Body	chr6:29796985	0.74	0.4	9.65E-03
HLA-G	TSS1500	chr6:29794008-29793607	0.876	0.856	2.97E-02
HLA-DPB1	Body	chr6:33053889	0.838	0.69	1.06E-02
ATF7IP	5'UTR	chr12:14522829	0.898	0.847	1.23E-02
AP2A2	3'UTR	chr11:1012211	0.994	0.987	1.49E-02
AP2A2	SSHORE	chr11:943737-944142	0.958	0.928	4.54E-02
WHSC1	Body	chr4:1977395	0.967	0.944	1.79E-02
CACNA1C	Body	chr12:2273119	0.994	0.986	2.67E-02
C10orf137	TSS200	chr10:127408000	0.0271	0.0167	3.36E-02
DYNC1H1	Body	chr14:102484867	0.991	0.988	3.45E-02
KIAA1217	Body	chr10:24792697	0.107	0.138	3.65E-02
KIAA1217	Body	chr10:24832948	0.962	0.944	3.65E-02
FBXL7	TSS200	chr5:15500249	0.0444	0.0319	4.28E-02
C19orf70/HSD11B1L	TSS200	chr19:5680956	0.0254	0.0138	4.74E-02

COL11A2	Body	chr6:33131893	0.657	0.755	1.45E-05
ASRGL1	Body	chr11:62123922	0.931	0.952	1.23E-02
CD3EAP/PPP1R13L	TSS1500	chr19:45908324	0.0664	0.073	1.47E-02
KIF26B	Body	chr1:245376028	0.918	0.942	1.54E-02
GLT1D1	Body	chr12:129467222	0.918	0.981	2.18E-02
LYN	5'UTR	chr8:56852253	0.126	0.149	2.72E-02
NCOR2	Body	chr12:124850781	0.98	0.993	2.81E-02
CACNA2D3	Body	chr3:54158728	0.799	0.88	3.01E-02
SCRN1	5'UTR	chr7:30028281	0.0552	0.0897	3.02E-02
KCNQ1	Body	chr11:2645098	0.926	0.945	3.59E-02
BRD2	5'UTR	chr6:32938950	0.0517	0.0896	3.98E-02
TRIM26	Body	chr6:30156254	0.958	0.973	4.09E-02
DIP2C	Body	chr10:665393	0.933	0.913	4.06E-02
DIP2C	Body	chr10:729479	0.97	0.937	5.82E-04
DIP2C	Body	chr10:557653	0.806	0.852	4.61E-02
MTUS2	Island	chr13:29913886-29914301	0.817	0.725	8.55E-04
MTUS2	NSHELF	chr13:29913886-29914301	0.703	0.617	1.95E-02

Annotated region:

- TSS1500 = 1500 bp upstream of the Transcription Start Site (TSS)
- TSS200 = 200 bp upstream of the TSS
- SSHORE = South shore of the CpG island. Shores are regions flanking island.
- NSHELF = North shelf of the CpG island. Shelves are regions flanking shores.

Genomic location: The unique genomic location of significant CpG site was indicated. The genomic location spanned by the CpG sites was indicated for significant region including several CpG sites.