Supplemental Material

Supplemental Tables

reads		counts
sequenced (PE1+PE2)		28,607,102
aligned		10,551,289
	uniquely aligned	5,656,887
	multiply aligned	4,894,402
not aligned		18,055,813
(aligned pairs		3,269,861)

Table S1. Summary of sequencing and alignment.

chromosome	unique	multiple, primary	multiple, secondary	CT strand unique+multiple	GA strand unique+multiple	ambiguous unique+multiple	inconsistent unique+multiple
chr2L	777,176	300,491	473,536	704,331	778,144	55,110	13,618
chr2LHet	2,353	4,484	8,702	6,973	7,557	1,007	2
chr2R	918,103	376,437	590,496	888,397	924,331	58,060	14,248
chr2RHet	22,844	44,843	86,204	73,632	72,494	7,691	74
chr3L	895,666	380,247	608,774	844,929	924,050	97,564	18,144
chr3LHet	20,230	41,584	79,222	69,894	66,991	4,107	44
chr3R	1,128,338	481,935	770,886	1,095,641	1,137,840	121,472	26,206
chr3RHet	20,661	38,654	71,630	60,010	62,820	8,063	52
chr4	12,482	17,669	32,282	22,702	35,870	3,847	14
chrX	1,783,287	1,247,351	1,812,379	2,440,229	2,113,552	263,102	26,134
chrXHet	2,266	4,051	7,953	6,524	7,311	433	2
chrYHet	1,232	3,472	6,571	5,522	5,162	591	0
chrU	72,236	1,953,063	4,036,585	458,907	993,522	4,607,289	2,166
chrM	13	121	247				
TOTAL	5,656,887	4,894,402	8,585,467				
TOTAL (-chrM)	5,656,874	4,894,281	8,585,220	6,677,691	7,129,644	5,228,336	100,704

Table S2. Alignments by chromosome.

Paired-read alignments are assigned to either the CT of the GA strand; only reads mapping to a strand are retained. If a read can be mapped with equal probability to either strand, it is labeled “ambiguous”. If the two reads of a pair do not map to the same strand, they are labeled “inconsistent”. Ambiguous and inconsistent reads are not analyzed further.

		methylated regions		resequenced regions	
Condition		CT strand	GA strand	CT strand	GA strand
2 methylated reads/position 3 methylated positions/region	positive	879	953	40	17
	negative	321	268	2	1
	undetermined	11337	11650	5	1
2 methylated reads/position 7 methylated positions/region	positive	383	388	31	11
	negative	215	196	2	1
	undetermined	11939	12287	14	7
	total number of regions	12537	12871	47	19

Table S3. Comparison with whole genome bisulfite data.
Methylated regions identified in this study were compared with data generated by whole genome bisulfite sequencing by Raddatz *et al.* "Positive" regions are methylated regions identified in this study that have supporting evidence in the Raddatz *et al.* data; we illustrate two conditions of different stringency (see Methods). "Negative" regions are regions identified in this study that have sufficient coverage in the Raddatz data to reveal methylation if present, but lack support in that data. "Undetermined" regions are not positive, and lack sufficient coverage to provide confidence that they are truly negative. The vast bulk of our methylated regions are positive or undetermined in the Raddatz data, and when positive they usually remain positive under the more stringent condition used to determine the status of a region. The coverage threshold we used to call regions as negative (100x) would allow detection of some regions methylated on ~1% of alleles. Many of the validated regions found in our study are methylated on <5% of alleles (Figure 3 and Figure S5). **Methods for comparison with whole-genome bisulfite sequencing data from Zemach *et al.* and Raddatz *et al.*:** The whole-genome bisulfite sequencing data for Stage 5 *Drosophila* embryo generated by Zemach *et al.* (Zemach et al. 2010) (accession #: GSM497255) and Raddatz *et al.* (Raddatz et al. 2013) (accession #: GSM983094) were downloaded from the GEO database. Sequence reads were aligned with Novoalign v2.07.11 with default options in bisulfite mode with the "b2" (directional) option and reporting only unique alignments, against a reference consisting of either the sequences that we validated by bisulfite PCR or the sequences of the 25,497 methylated regions identified in this study. Reads were assigned to the CT or GA strands according to Novoalign mappings. Reads aligning to the strand that was not amplified were discarded, and reads that are potential PCR duplicates were removed with MarkDuplicates from the Picard suite (picard.sourceforge.net). We determined sequence coverage and the percentage of methylation at each cytosine using the output of mpileup from the samtools suite (Li et al. 2009). We define a cytosine as methylated if at least two reads are unconverted at that position. To evaluate the agreement between Raddatz *et al.* and our data, the reference regions described above were divided into three groups according to the evidence for methylation in Raddatz *et al.*'s data: "positive", if the region contains at least three methylated cytosines (or seven in the more stringent condition); "negative", if the region contains less than 3 methylated cytosines and at least three (or seven) cytosines with coverage greater than 100 reads; "undetermined", if the region is neither positive nor negative.

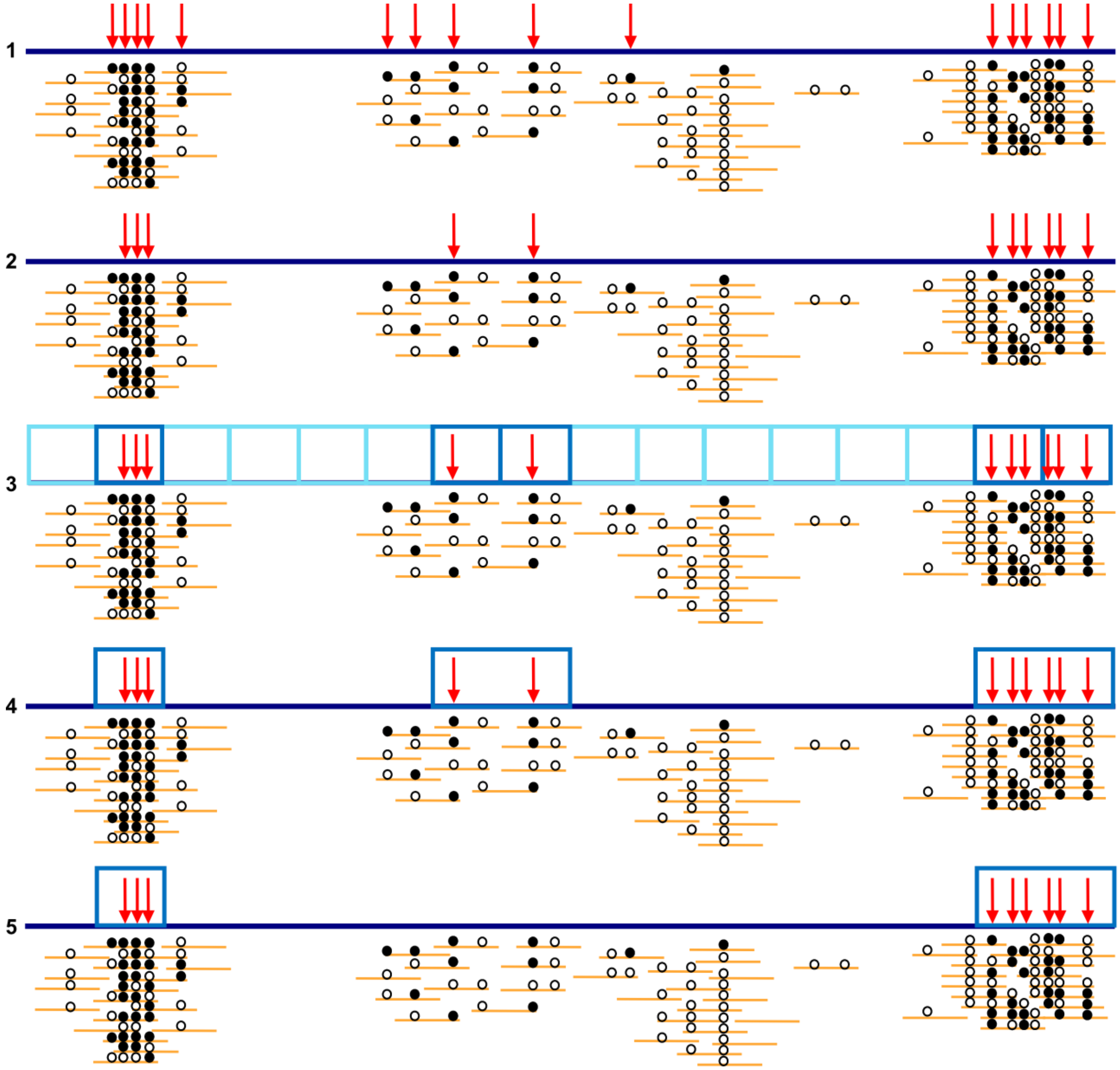


Figure S1. Methylated regions in the genome of Stage 5 *Drosophila* embryos. Procedure for identification of methylated cytosines and methylated regions from MeDIP-Bseq data. The steps illustrated here led to the identification of the 25,497 methylated regions discussed in the text. Red arrows identify cytosine positions that pass the filter at each step; the steps are applied sequentially to the output of the preceding step.

- Step 1 identifies cytosine positions at which the ratio of C-containing alignments (unconverted, i.e. methylated) over the sum of C- and T-containing alignments was greater than 0.1 (methylated cytosine: closed circles; unmethylated: open circles).
- Step 2 identifies cytosine positions at which at least three alignments contain a methylated cytosine.
- Step 3 divides the genome into contiguous 25-base segments.
- Step 4 removes segments that do not contain cytosines passing Step 2, and merges contiguous segments.
- Step 5 retains only those segments in which the alignments contained at least 25 methylated cytosines; these are the methylated regions.

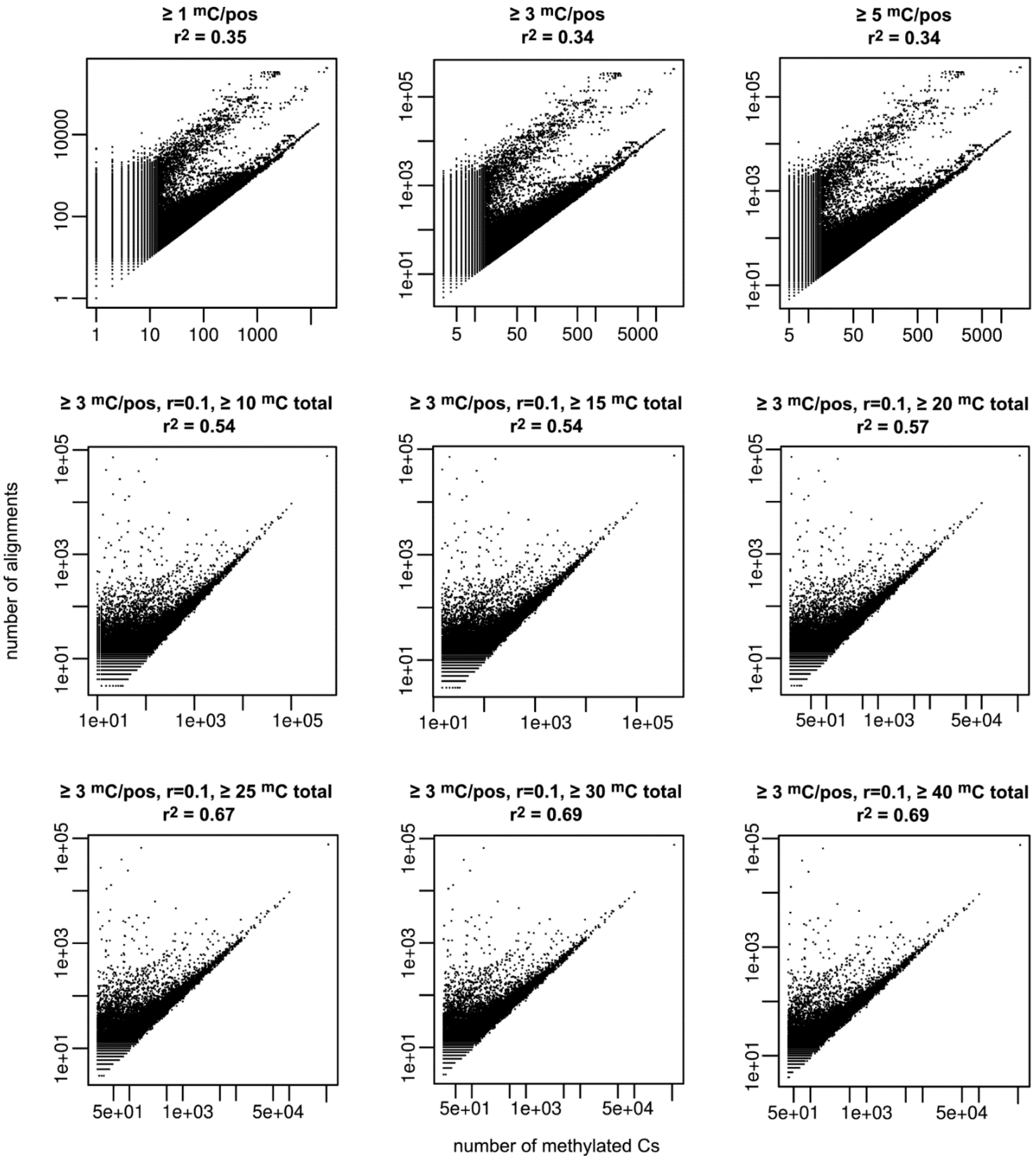


Figure S2. Effects of different parameter choices on the identification of methylated cytosines in the MeDIP-Bseq data. In all of the plots displayed, the number of methylated cytosines in the reads aligning to a given cytosine in the reference sequence is shown on the x-axis, and the number of sequence reads containing that position is shown on the y-axis. Only those cytosines meeting the condition shown at the top of each plot are displayed (see Methods for a detailed description of the conditions). The procedure described in the text and Figure S1 removes cytosine positions with weakly supported methylation states. The top row is derived from step 2 in Figure S1: it illustrates the effect of requiring more methylated reads supporting the status of the position. The middle and bottom rows are derived from the step shown in Figure S1, step 5: only those cytosines included in one of the regions meeting the parameters are shown. For each set of parameters (denoted at the top of each plot), the correlation coefficient between the number of methylated cytosines and the number of alignments within a region was calculated. We chose the set of parameters that optimizes the correlation coefficient at the lowest cost in discarded methylated regions.

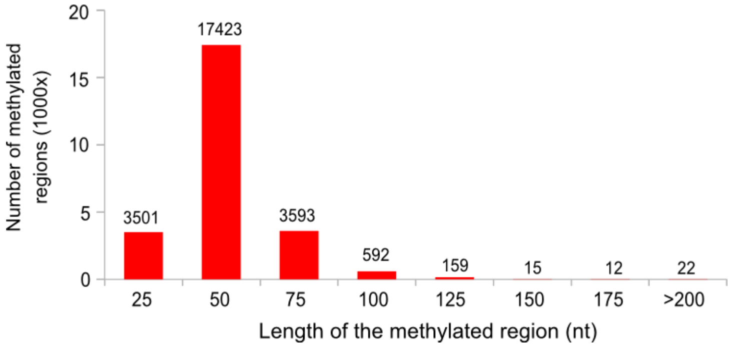


Figure S3. Length distribution of methylated regions. The number of methylated regions of a given length range is shown at the top of each column. 97% of the regions have a length of 75 bases or less.

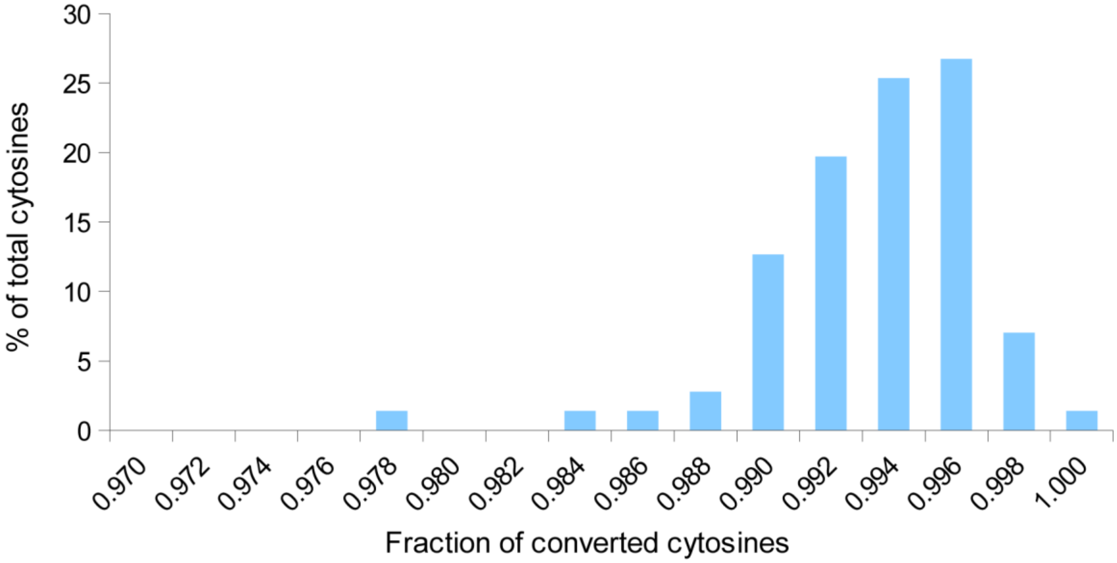
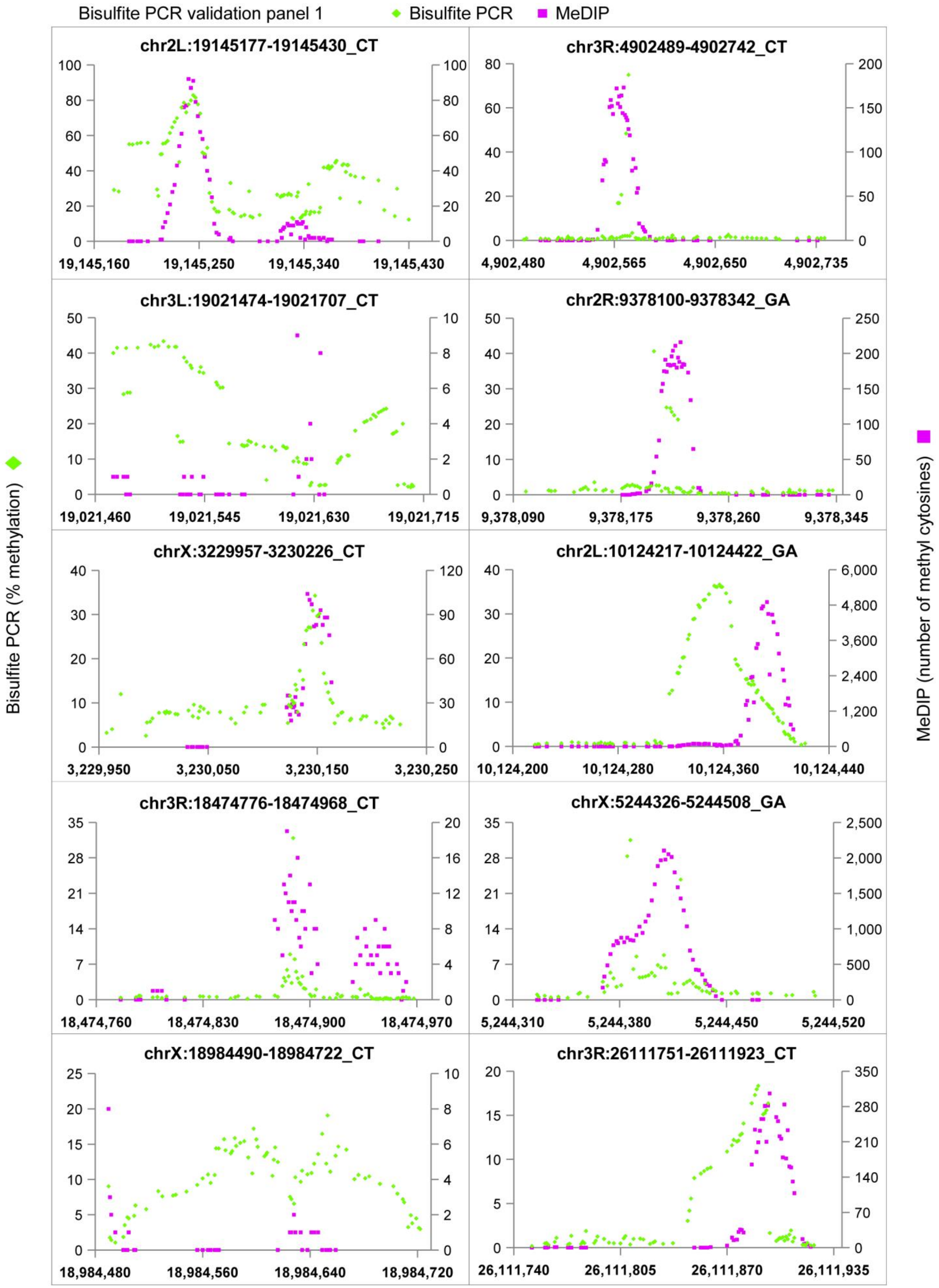
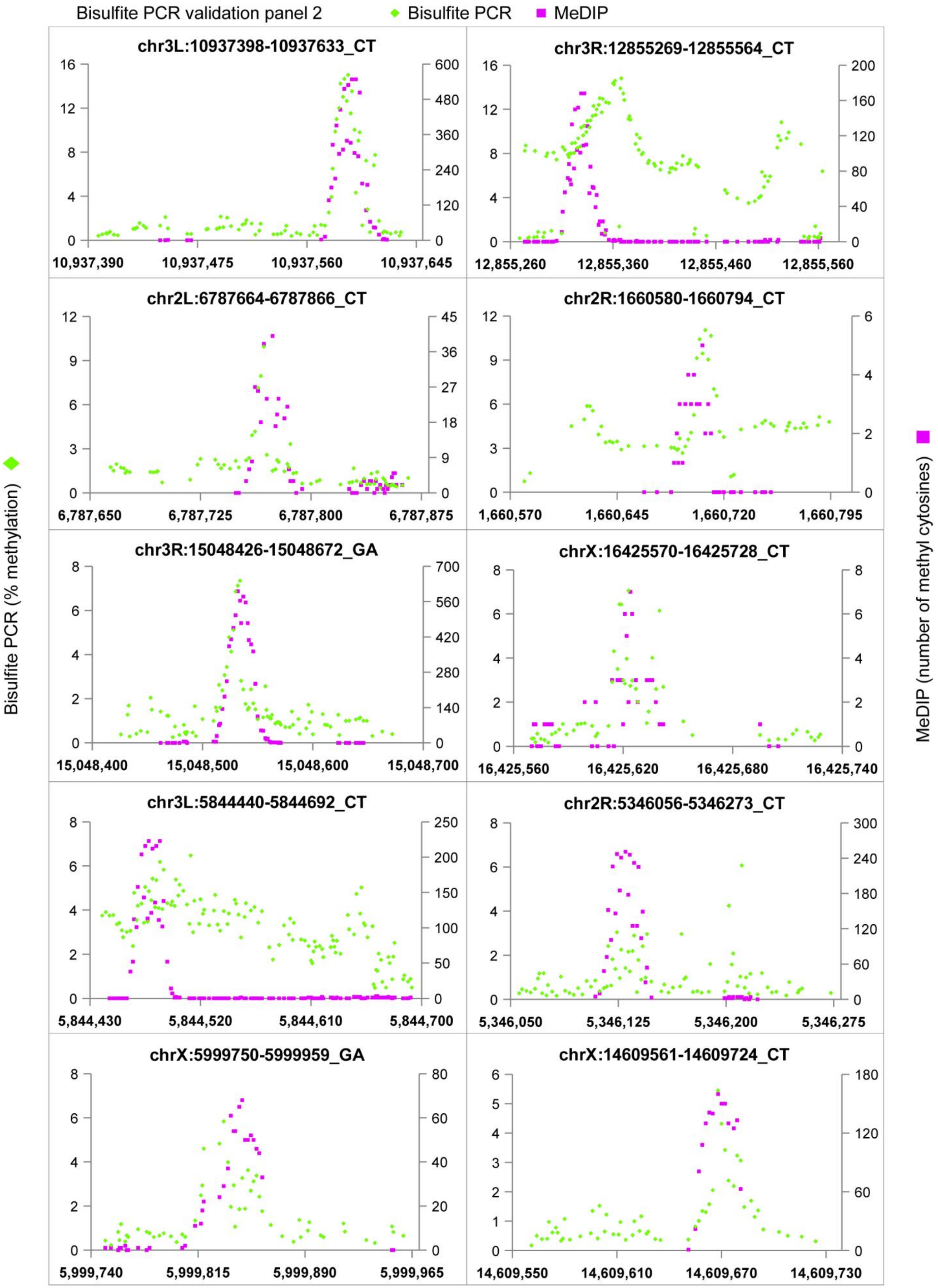
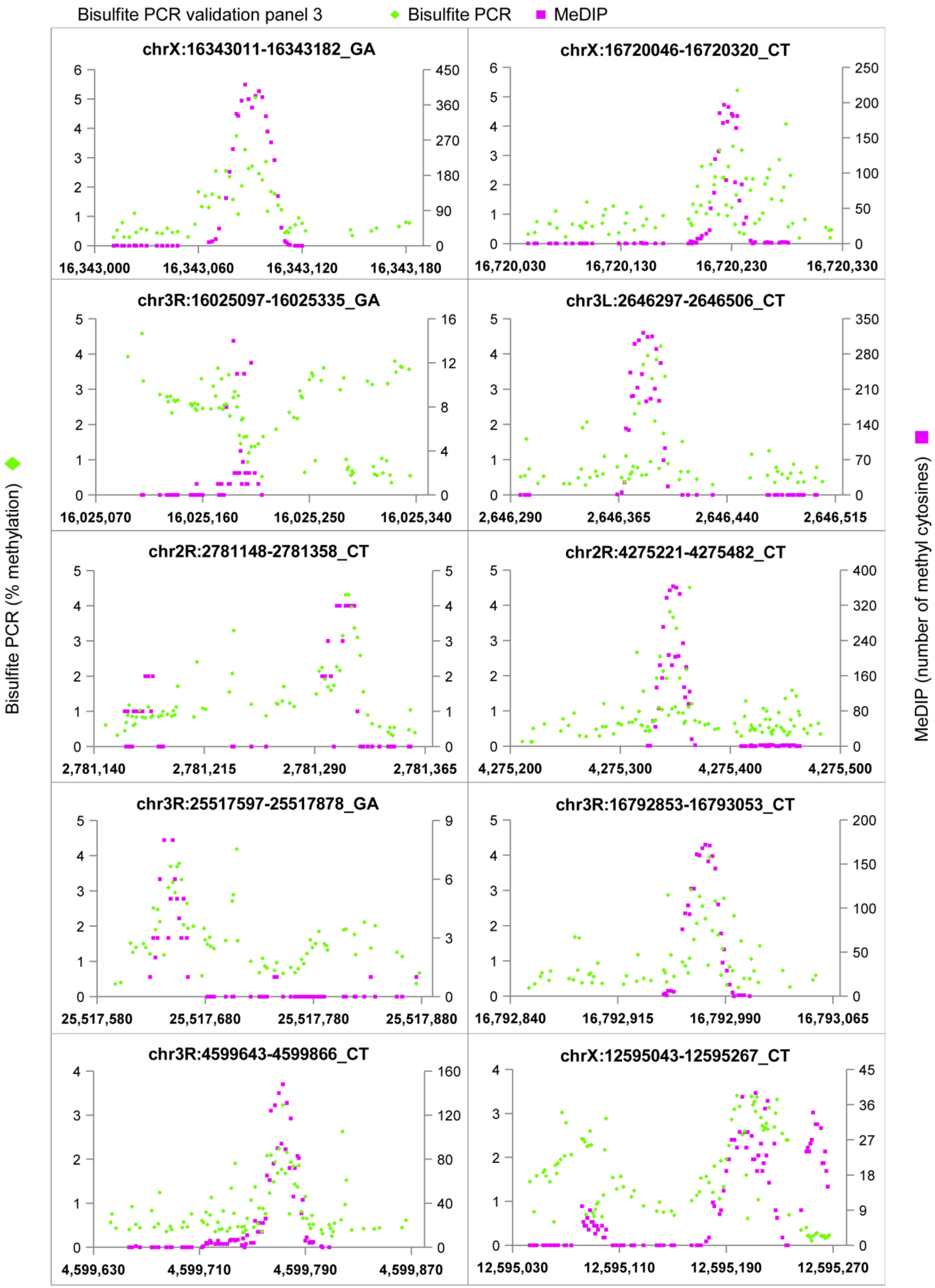
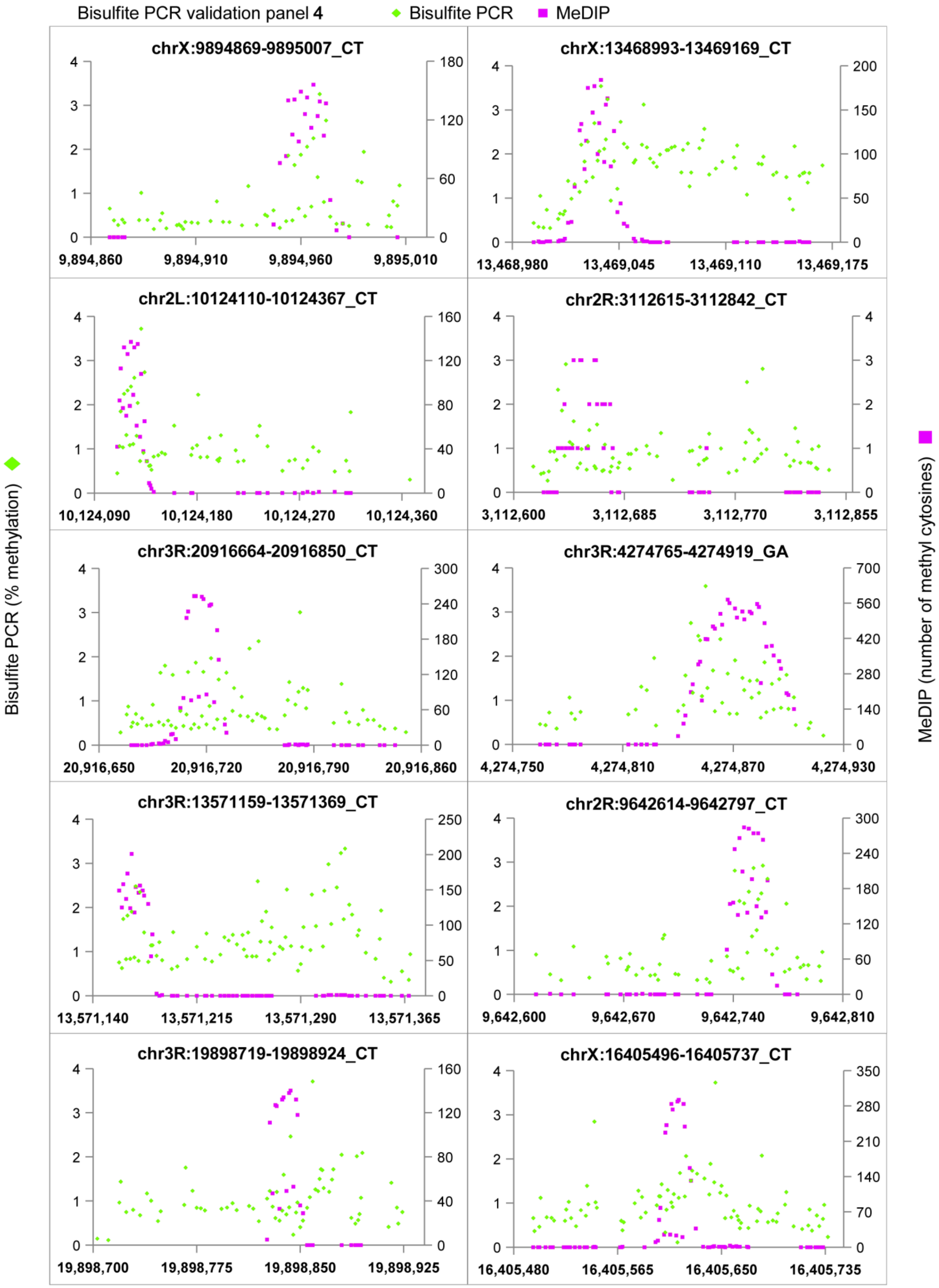


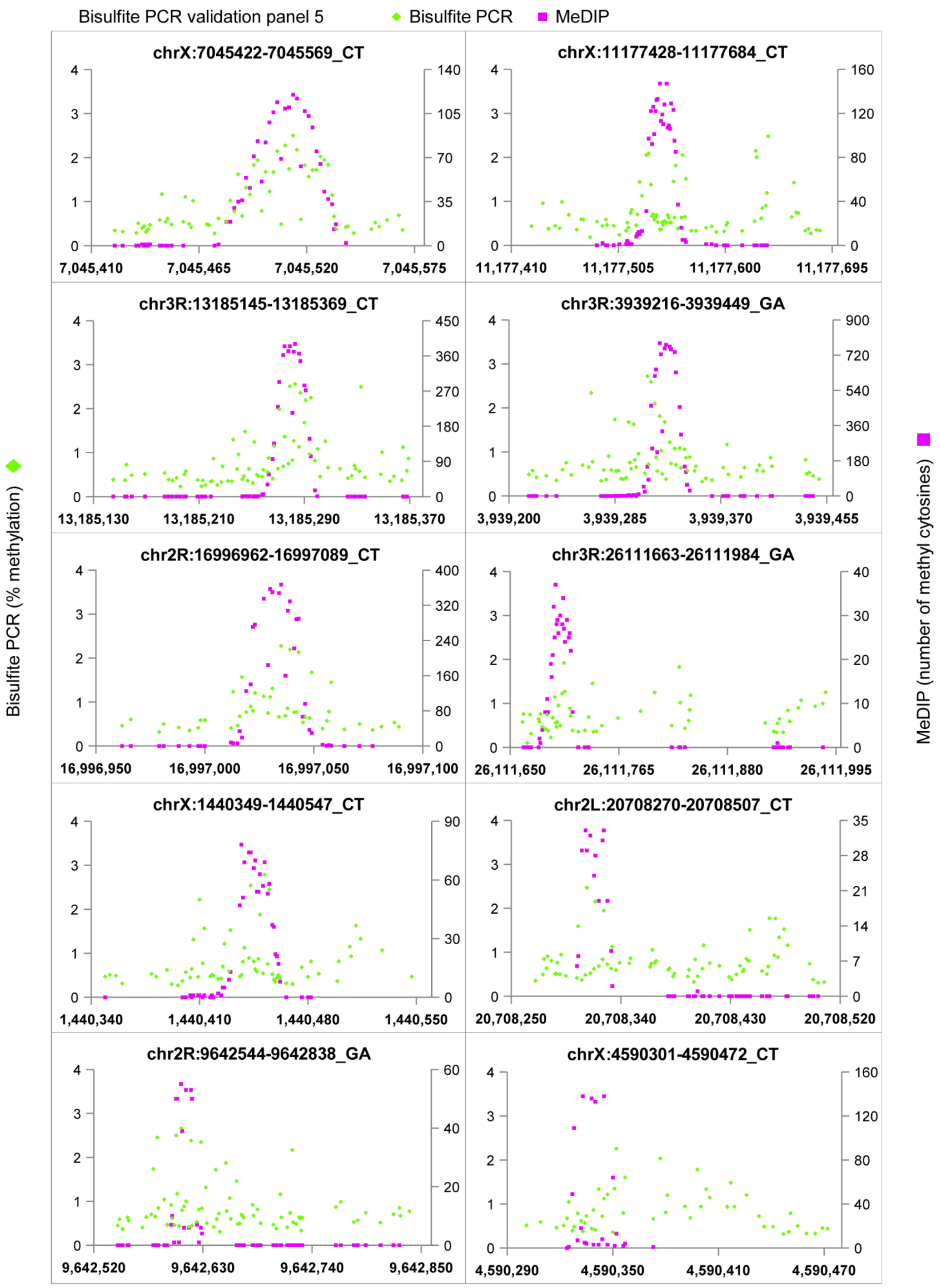
Figure S4. Efficiency of bisulfite conversion as determined by bisulfite-PCR sequencing of lambda phage DNA. Lambda phage DNA, grown in an *E. coli* strain deficient for methylation, was bisulfite converted in the same reaction as the *Drosophila* DNA used for the validation in Figures 3, S5, S6 and S8. Two segments of the lambda genome were amplified and sequenced to a median coverage of 109, 207 reads (range 82,602-161,273). The x-axis shows the rate of conversion from C to T, determined as the ration of T over C +T at each cytosine position. The y-axis shows the percent of total cytosines with a given conversion rate. Average conversion is >99%, and 96% of all cytosine positions had a conversion rate ≥ 0.988.

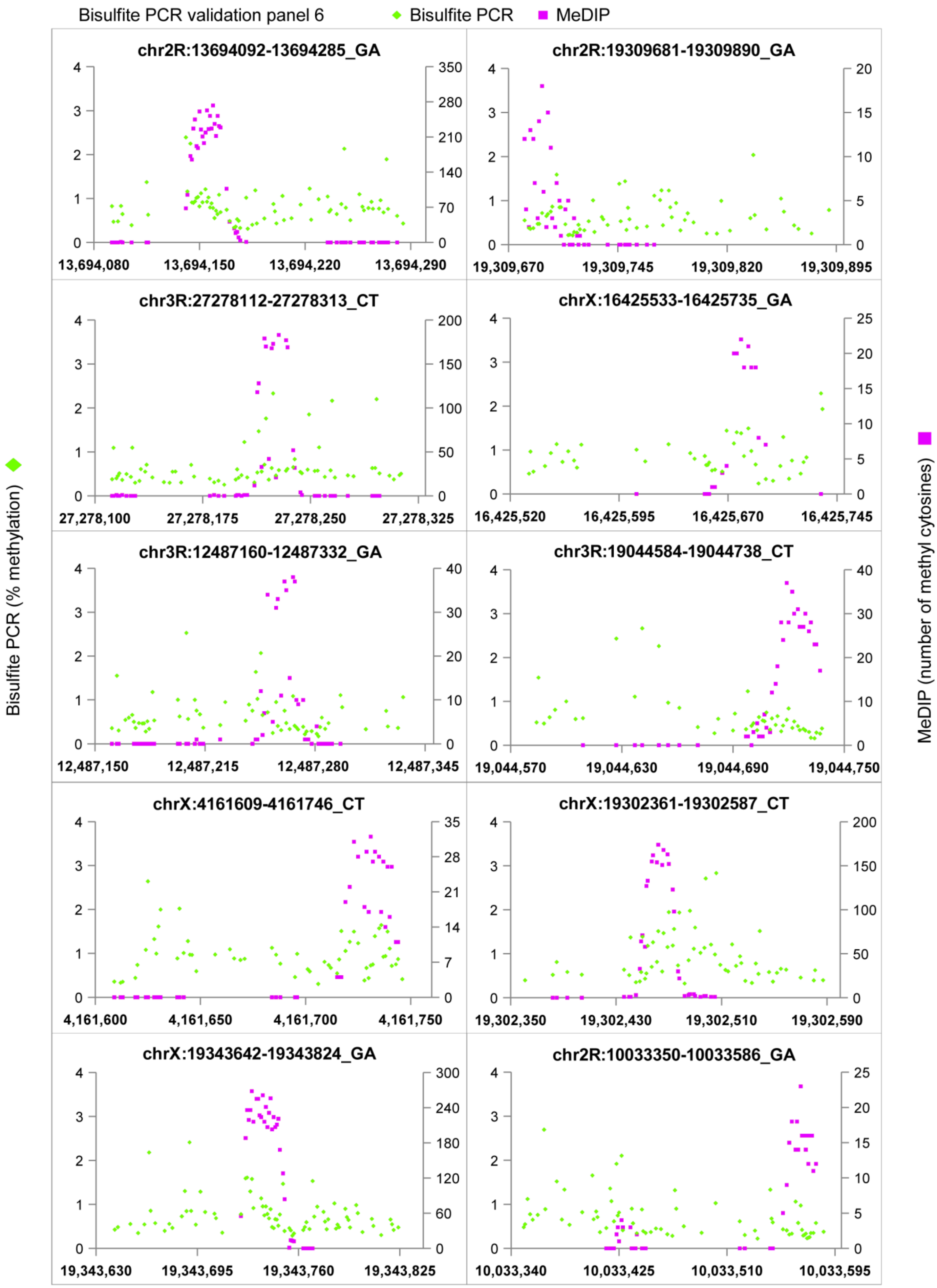

Bisulfite PCR validation panel 1
◆ Bisulfite PCR
■ MeDIP
Bisulfite PCR (% methylation)
MeDIP (number of methyl cytosines)
chr2L:19145177-19145430_CT
0 20 40 60 80 100
19,145,160 19,145,250 19,145,340 19,145,430
chr3R:4902489-4902742_CT
0 20 40 60 80
0 50 100 150 200
4,902,480 4,902,565 4,902,650 4,902,735
chr3L:19021474-19021707_CT
0 10 20 30 40 50
0 2 4 6 8 10
19,021,460 19,021,545 19,021,630 19,021,715
chr2R:9378100-9378342_GA
0 10 20 30 40 50
0 50 100 150 200 250
9,378,090 9,378,175 9,378,260 9,378,345
chrX:3229957-3230226_CT
0 10 20 30 40
0 30 60 90 120
3,229,950 3,230,050 3,230,150 3,230,250
chr2L:10124217-10124422_GA
0 10 20 30 40
0 1,200 2,400 3,600 4,800 6,000
10,124,200 10,124,280 10,124,360 10,124,440
chr3R:18474776-18474968_CT
0 7 14 21 28 35
0 4 8 12 16 20
18,474,760 18,474,830 18,474,900 18,474,970
chrX:5244326-5244508_GA
0 7 14 21 28 35
0 500 1,000 1,500 2,000 2,500
5,244,310 5,244,380 5,244,450 5,244,520
chrX:18984490-18984722_CT
0 5 10 15 20 25
0 2 4 6 8 10
18,984,480 18,984,560 18,984,640 18,984,720
chr3R:26111751-26111923_CT
0 5 10 15 20
0 70 140 210 280 350
26,111,740 26,111,805 26,111,870 26,111,935


Bisulfite PCR validation panel 2
◆ Bisulfite PCR
■ MeDIP
Bisulfite PCR (% methylation) ◆
MeDIP (number of methyl cytosines) ■
chr3L:10937398-10937633_CT
0 4 8 12 16
0 120 240 360 480 600
10,937,390 10,937,475 10,937,560 10,937,645
chr3R:12855269-12855564_CT
0 4 8 12 16
0 40 80 120 160 200
12,855,260 12,855,360 12,855,460 12,855,560
chr2L:6787664-6787866_CT
0 3 6 9 12
0 9 18 27 36 45
6,787,650 6,787,725 6,787,800 6,787,875
chr2R:1660580-1660794_CT
0 3 6 9 12
0 2 4 6
1,660,570 1,660,645 1,660,720 1,660,795
chr3R:15048426-15048672_GA
0 2 4 6 8
0 140 280 420 560 700
15,048,400 15,048,500 15,048,600 15,048,700
chrX:16425570-16425728_CT
0 2 4 6 8
0 2 4 6 8
16,425,560 16,425,620 16,425,680 16,425,740
chr3L:5844440-5844692_CT
0 2 4 6 8
0 50 100 150 200 250
5,844,430 5,844,520 5,844,610 5,844,700
chr2R:5346056-5346273_CT
0 2 4 6 8
0 60 120 180 240 300
5,346,050 5,346,125 5,346,200 5,346,275
chrX:5999750-5999959_GA
0 2 4 6 8
0 20 40 60 80
5,999,740 5,999,815 5,999,890 5,999,965
chrX:14609561-14609724_CT
0 1 2 3 4 5 6
0 60 120 180
14,609,550 14,609,610 14,609,670 14,609,730


Bisulfite PCR validation panel 3
Bisulfite PCR
MeDIP
Bisulfite PCR (% methylation)
MeDIP (number of methyl cytosines)
chrX:16343011-16343182_GA
0 1 2 3 4 5 6
0 90 180 270 360 450
16,343,000 16,343,060 16,343,120 16,343,180
chrX:16720046-16720320_CT
0 1 2 3 4 5 6
0 50 100 150 200 250
16,720,030 16,720,130 16,720,230 16,720,330
chr3R:16025097-16025335_GA
0 1 2 3 4 5
0 4 8 12 16
16,025,070 16,025,160 16,025,250 16,025,340
chr3L:2646297-2646506_CT
0 1 2 3 4 5
0 70 140 210 280 350
2,646,290 2,646,365 2,646,440 2,646,515
chr2R:2781148-2781358_CT
0 1 2 3 4 5
0 1 2 3 4 5
2,781,140 2,781,215 2,781,290 2,781,365
chr2R:4275221-4275482_CT
0 1 2 3 4 5
0 80 160 240 320 400
4,275,200 4,275,300 4,275,400 4,275,500
chr3R:25517597-25517878_GA
0 1 2 3 4 5
0 3 6 9
25,517,580 25,517,680 25,517,780 25,517,880
chr3R:16792853-16793053_CT
0 1 2 3 4 5
0 50 100 150 200
16,792,840 16,792,915 16,792,990 16,793,065
chr3R:4599643-4599866_CT
0 1 2 3 4
0 40 80 120 160
4,599,630 4,599,710 4,599,790 4,599,870
chrX:12595043-12595267_CT
0 1 2 3 4
0 9 18 27 36 45
12,595,030 12,595,110 12,595,190 12,595,270


Bisulfite PCR validation panel 4
◆ Bisulfite PCR
■ MeDIP
Bisulfite PCR (% methylation) ◆
MeDIP (number of methyl cytosines) ■
chrX:9894869-9895007_CT
0 1 2 3 4
0 60 120 180
9,894,860 9,894,910 9,894,960 9,895,010
chrX:13468993-13469169_CT
0 1 2 3 4
0 50 100 150 200
13,468,980 13,469,045 13,469,110 13,469,175
chr2L:10124110-10124367_CT
0 1 2 3 4
0 40 80 120 160
10,124,090 10,124,180 10,124,270 10,124,360
chr2R:3112615-3112842_CT
0 1 2 3 4
0 1 2 3 4
3,112,600 3,112,685 3,112,770 3,112,855
chr3R:20916664-20916850_CT
0 1 2 3 4
0 60 120 180 240 300
20,916,650 20,916,720 20,916,790 20,916,860
chr3R:4274765-4274919_GA
0 1 2 3 4
0 140 280 420 560 700
4,274,750 4,274,810 4,274,870 4,274,930
chr3R:13571159-13571369_CT
0 1 2 3 4
0 50 100 150 200 250
13,571,140 13,571,215 13,571,290 13,571,365
chr2R:9642614-9642797_CT
0 1 2 3 4
0 60 120 180 240 300
9,642,600 9,642,670 9,642,740 9,642,810
chr3R:19898719-19898924_CT
0 1 2 3 4
0 40 80 120 160
19,898,700 19,898,775 19,898,850 19,898,925
chrX:16405496-16405737_CT
0 1 2 3 4
0 70 140 210 280 350
16,405,480 16,405,565 16,405,650 16,405,735


Bisulfite PCR validation panel 5
◆ Bisulfite PCR
■ MeDIP
Bisulfite PCR (% methylation) ◆
MeDIP (number of methyl cytosines) ■
chrX:7045422-7045569_CT
0 1 2 3 4
0 35 70 105 140
7,045,410 7,045,465 7,045,520 7,045,575
chrX:11177428-11177684_CT
0 1 2 3 4
0 40 80 120 160
11,177,410 11,177,505 11,177,600 11,177,695
chr3R:13185145-13185369_CT
0 1 2 3 4
0 90 180 270 360 450
13,185,130 13,185,210 13,185,290 13,185,370
chr3R:3939216-3939449_GA
0 1 2 3 4
0 180 360 540 720 900
3,939,200 3,939,285 3,939,370 3,939,455
chr2R:16996962-16997089_CT
0 1 2 3 4
0 80 160 240 320 400
16,996,950 16,997,000 16,997,050 16,997,100
chr3R:26111663-26111984_GA
0 1 2 3 4
0 10 20 30 40
26,111,650 26,111,765 26,111,880 26,111,995
chrX:1440349-1440547_CT
0 1 2 3 4
0 30 60 90
1,440,340 1,440,410 1,440,480 1,440,550
chr2L:20708270-20708507_CT
0 1 2 3 4
0 7 14 21 28 35
20,708,250 20,708,340 20,708,430 20,708,520
chr2R:9642544-9642838_GA
0 1 2 3 4
0 20 40 60
9,642,520 9,642,630 9,642,740 9,642,850
chrX:4590301-4590472_CT
0 1 2 3 4
0 40 80 120 160
4,590,290 4,590,350 4,590,410 4,590,470


Bisulfite PCR validation panel 6
◆ Bisulfite PCR
■ MeDIP
Bisulfite PCR (% methylation)
MeDIP (number of methyl cytosines)
chr2R:13694092-13694285_GA
4
3
2
1
0
350
280
210
140
70
0
13,694,080
13,694,150
13,694,220
13,694,290
chr2R:19309681-19309890_GA
20
15
10
5
19,309,670
19,309,745
19,309,820
19,309,895
chr3R:27278112-27278313_CT
200
150
100
50
27,278,100
27,278,175
27,278,250
27,278,325
chrX:16425533-16425735_GA
25
20
15
10
5
16,425,520
16,425,595
16,425,670
16,425,745
chr3R:12487160-12487332_GA
40
30
20
10
12,487,150
12,487,215
12,487,280
12,487,345
chr3R:19044584-19044738_CT
40
30
20
10
19,044,570
19,044,630
19,044,690
19,044,750
chrX:4161609-4161746_CT
35
28
21
14
7
4,161,600
4,161,650
4,161,700
4,161,750
chrX:19302361-19302587_CT
200
150
100
50
19,302,350
19,302,430
19,302,510
19,302,590
chrX:19343642-19343824_GA
300
240
180
120
60
19,343,630
19,343,695
19,343,760
19,343,825
chr2R:10033350-10033586_GA
25
20
15
10
5
10,033,340
10,033,425
10,033,510
10,033,595

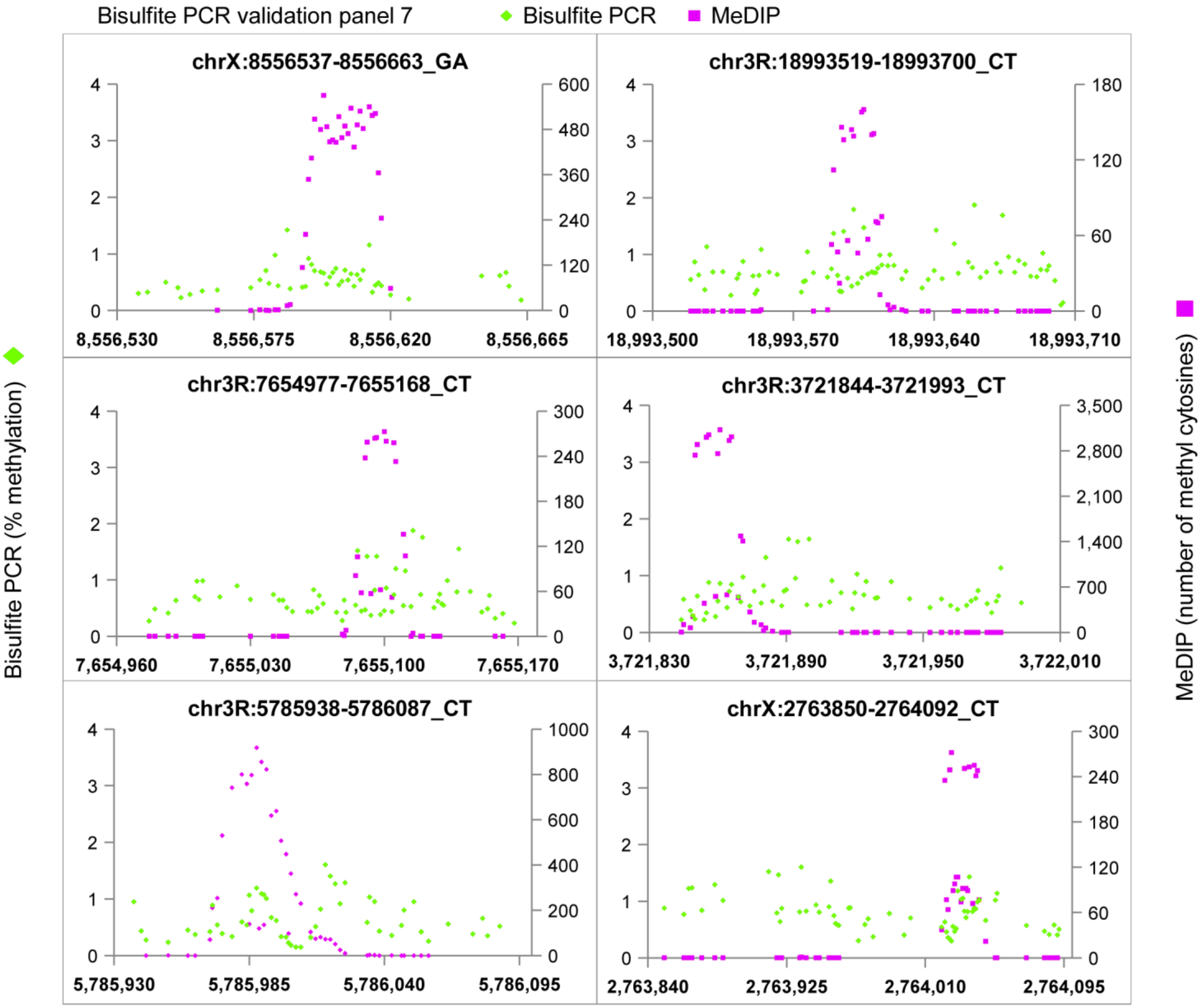


Figure S5. Direct amplification of bisulfite-converted DNA confirms methylation patterns. The full set of 66 regions that were analyzed is shown. Methylated regions identified by MeDIP-bisulfite sequencing were PCR amplified from bisulfite converted DNA and Illumina sequenced to at least 10,000X coverage. Each dot represents one cytosine (green – bisulfite PCR; purple – MeDIP bisulfite). The y-axis at the left indicates the percent of methylated cytosines in the bisulfite PCR; the y-axis at the right indicates the number of methylated cytosines detected by MeDIP bisulfite. While the MeDIP bisulfite analysis is not quantitative, bisulfite PCR demonstrates the proportion of methylated cytosines at a given position, as well as the pattern of methylation of the amplified region. There is good agreement in the pattern of methylation detected by the two methods.

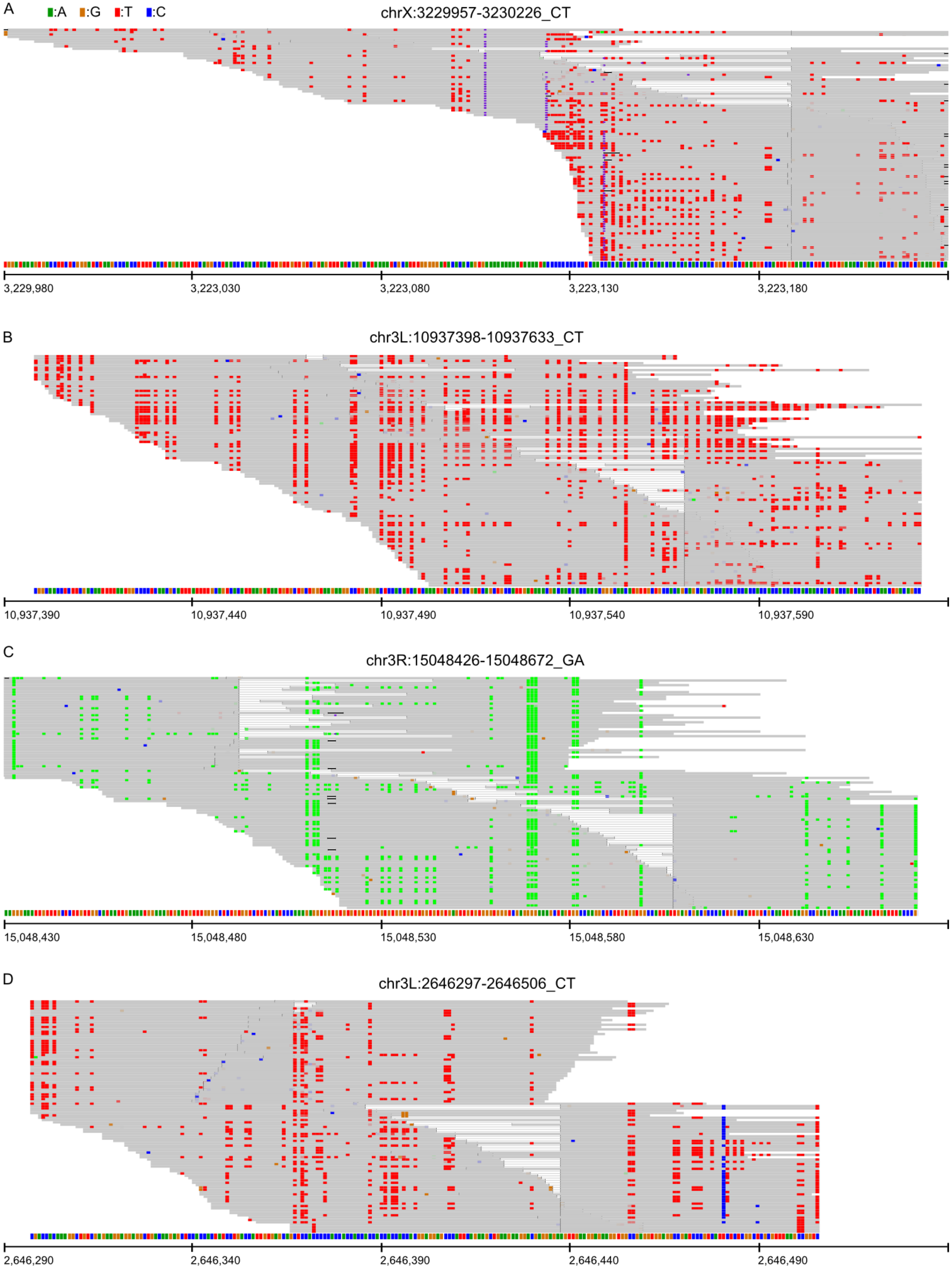
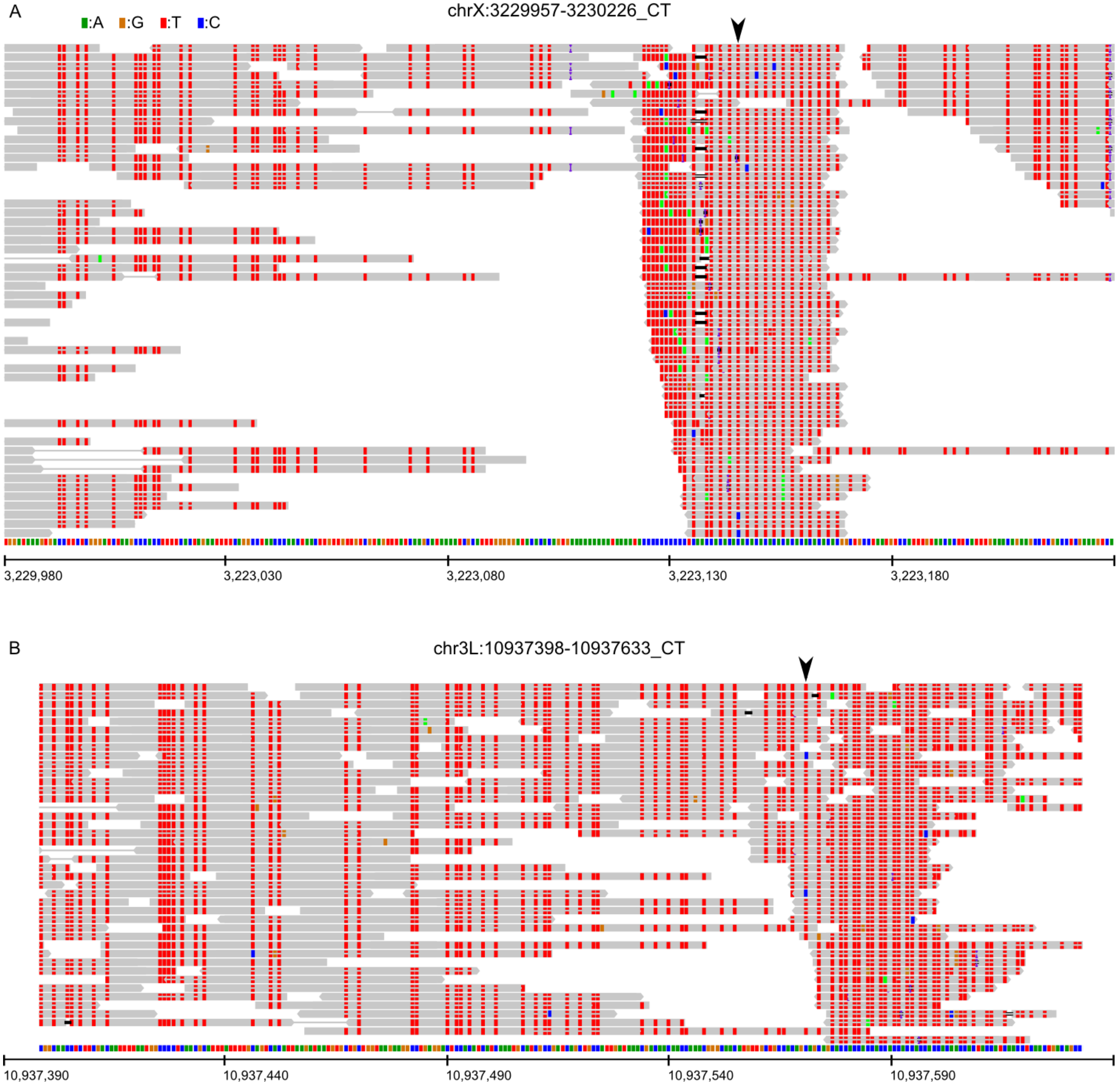

A
🟩:A 🟧:G 🟥:T 🟦:C
chrX:3229957-3230226_CT
3,229,980
3,223,030
3,223,080
3,223,130
3,223,180
B
chr3L:10937398-10937633_CT
10,937,390
10,937,440
10,937,490
10,937,540
10,937,590
C
chr3R:15048426-15048672_GA
15,048,430
15,048,480
15,048,530
15,048,580
15,048,630
D
chr3L:2646297-2646506_CT
2,646,290
2,646,340
2,646,390
2,646,440
2,646,490

Figure S6. Visualization of alignments at four methylated regions, illustrating correct alignment of reads that support cytosine methylation. Shown are 100 base (including a 6 base index) paired end reads aligning to the regions displayed in Figure 3; these reads derived from wild type *EP(2)GE 15695* flies. Only a subset of read pairs (100) is shown; reads were selected based on their content of unconverted cytosines, in order to illustrate the alignment of such reads. Each line shows the alignment of a read pair; when the two sequences of a pair of reads do not overlap, a thin grey line shows their connection. Alignments are displayed using the “collapsed” mode of the Integrative Genome Viewer; in this mode, the direction of the alignments is shown by a vertical grey line at the 3’ end of the alignment. The color-coded reference sequence is at the bottom of each panel, with the color key shown at the top of the figure. A match between a read and the reference is shown in grey; a mismatch is shown with the color of the mismatched base. Unmethylated cytosines are sequenced as ‘T’ (red) on the CT strand (panels A, B, and D) and as ‘A’ (green) on the GA strand (panel C); thus converted (unmethylated) cytosines are shown in color, and any methylated cytosines are denoted by gray color at a position that is colored in other reads. The figure shows that alignments of reads containing unconverted (methylated) cytosines are unambiguous and extend well beyond the low complexity sequences where methylation is concentrated.


A
🟩:A 🟧:G 🟥:T 🟦:C
chrX:3229957-3230226_CT
3,229,980
3,223,030
3,223,080
3,223,130
3,223,180
B
chr3L:10937398-10937633_CT
10,937,390
10,937,440
10,937,490
10,937,540
10,937,590

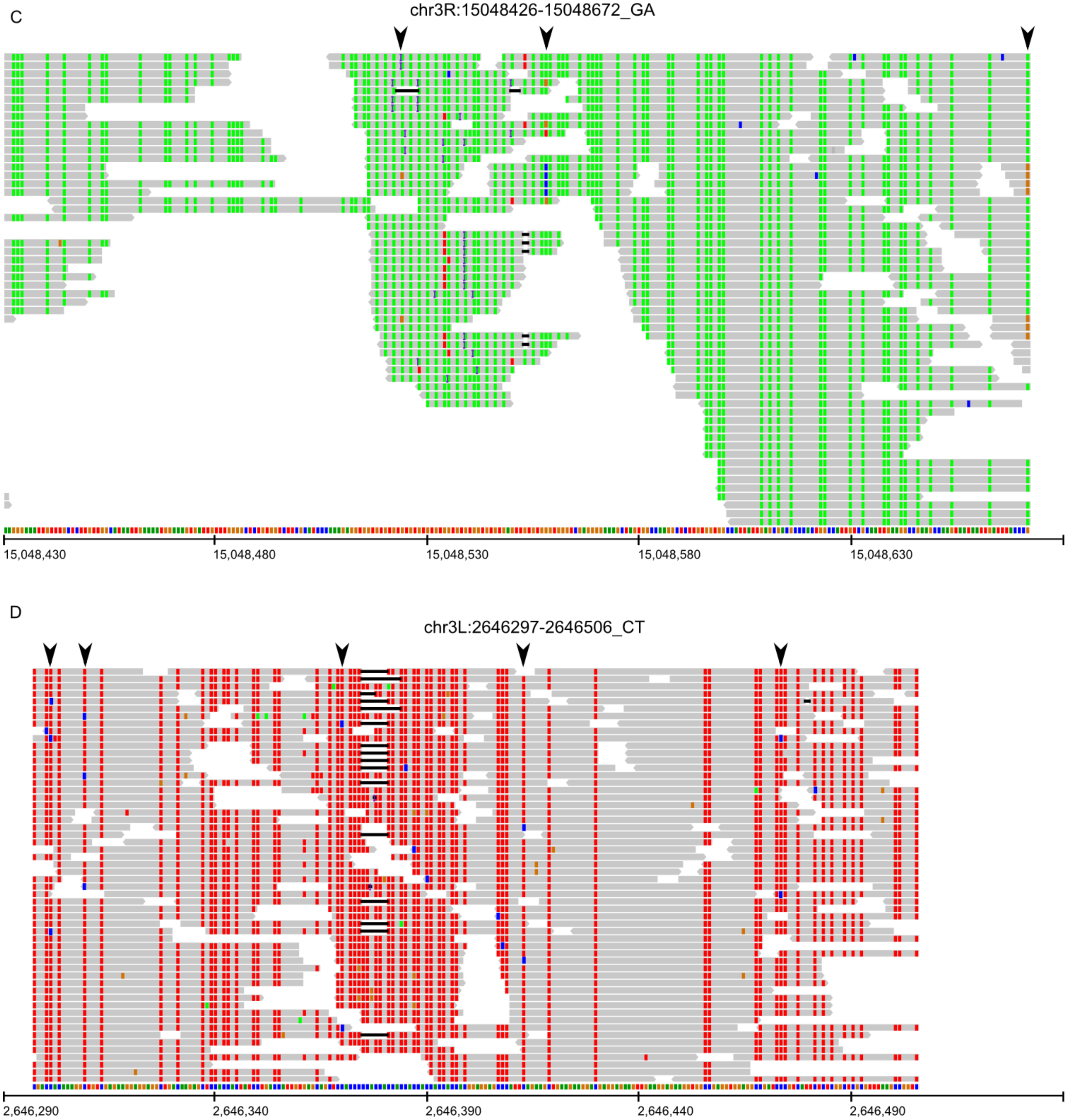
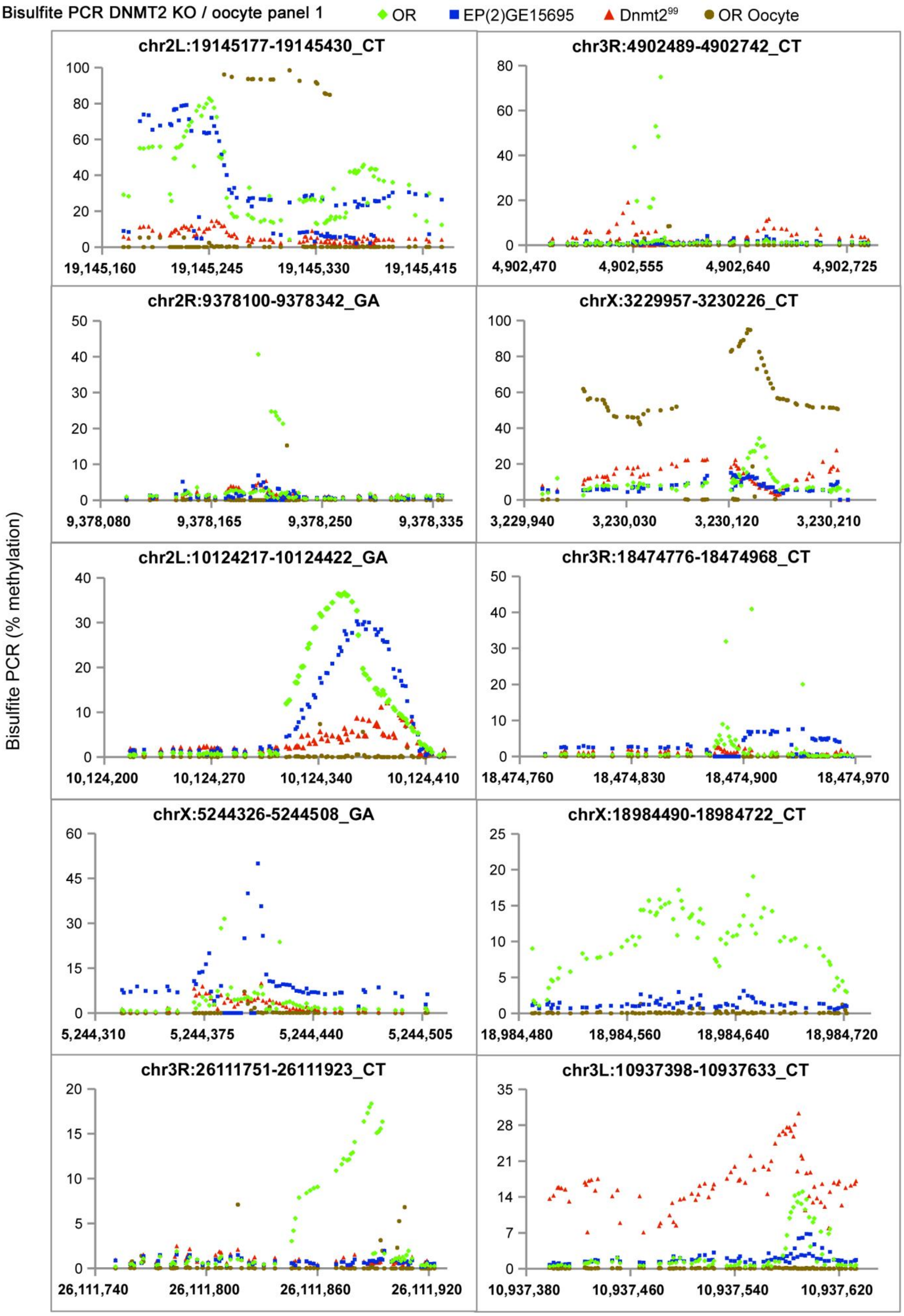
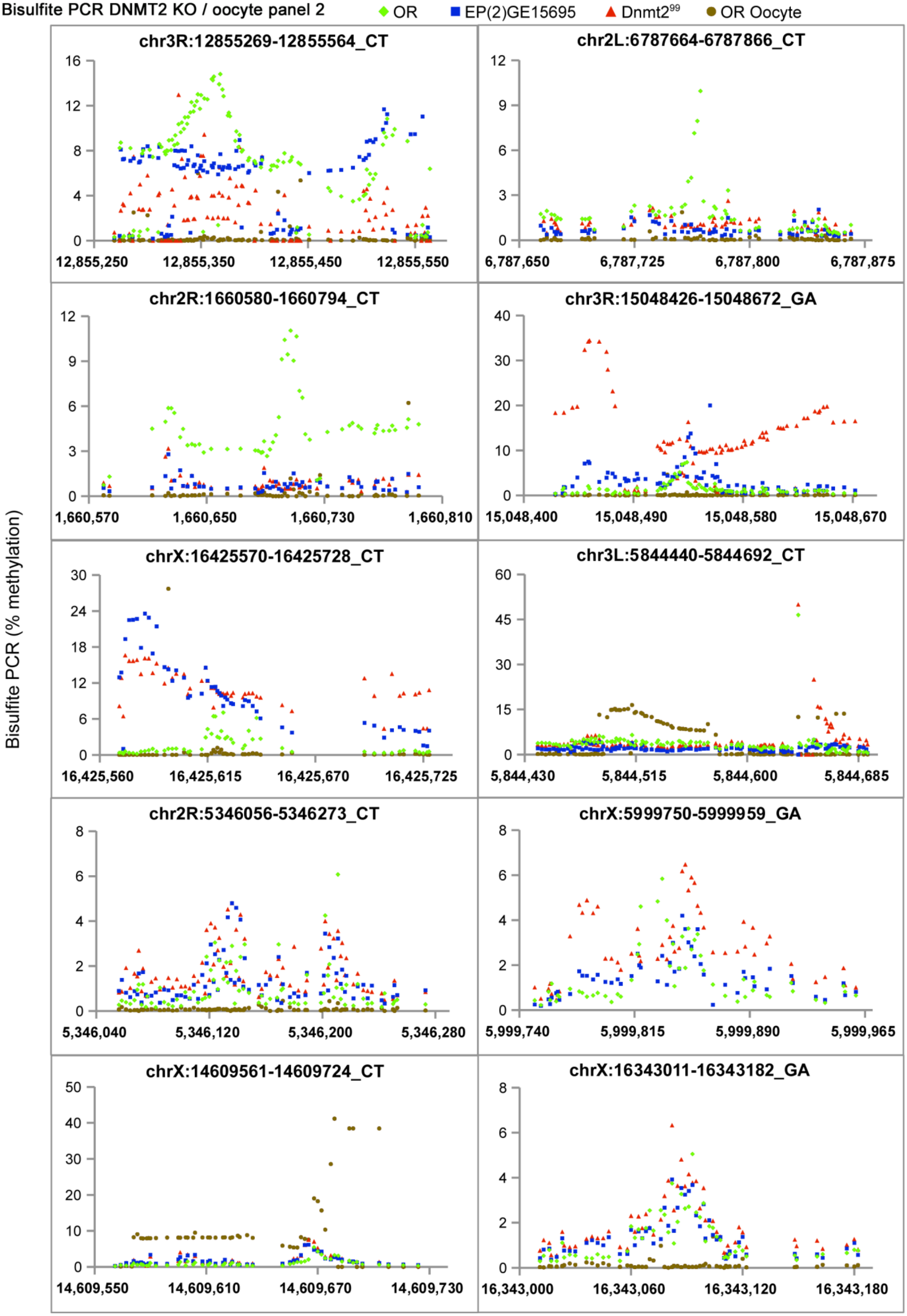


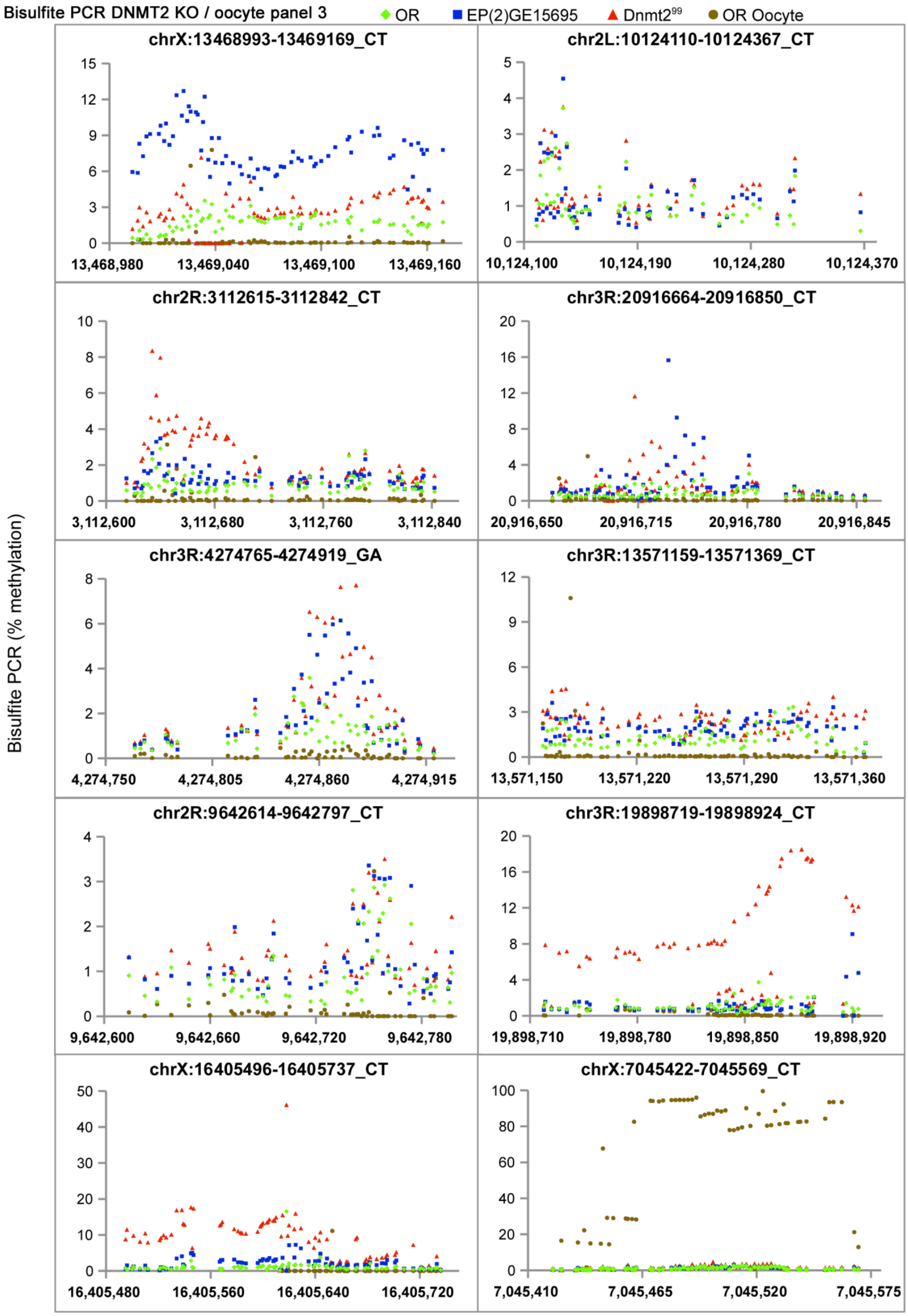
Figure S7. Reads from the dataset of Raddatz et al., aligned to the regions displayed in Figure 3. As in Fig. S6, each line shows the alignment of a read pair; when the two sequences of a pair of reads do not overlap, a thin grey line shows their connection. The color-coded reference sequence is at the bottom of each panel, with the color key shown at the top of the figure. A match between a read and the reference is shown in grey; a mismatch is shown with the color of the mismatched base. Unmethylated cytosines are sequenced as 'T' (red) on the CT strand (panels A, B, and D) and as 'A' (green) on the GA strand (panel C). Any methylated cytosines are denoted by blue color in A, B, and D, and in brown in C. Arrows mark positions that contain at least two unconverted cytosines.

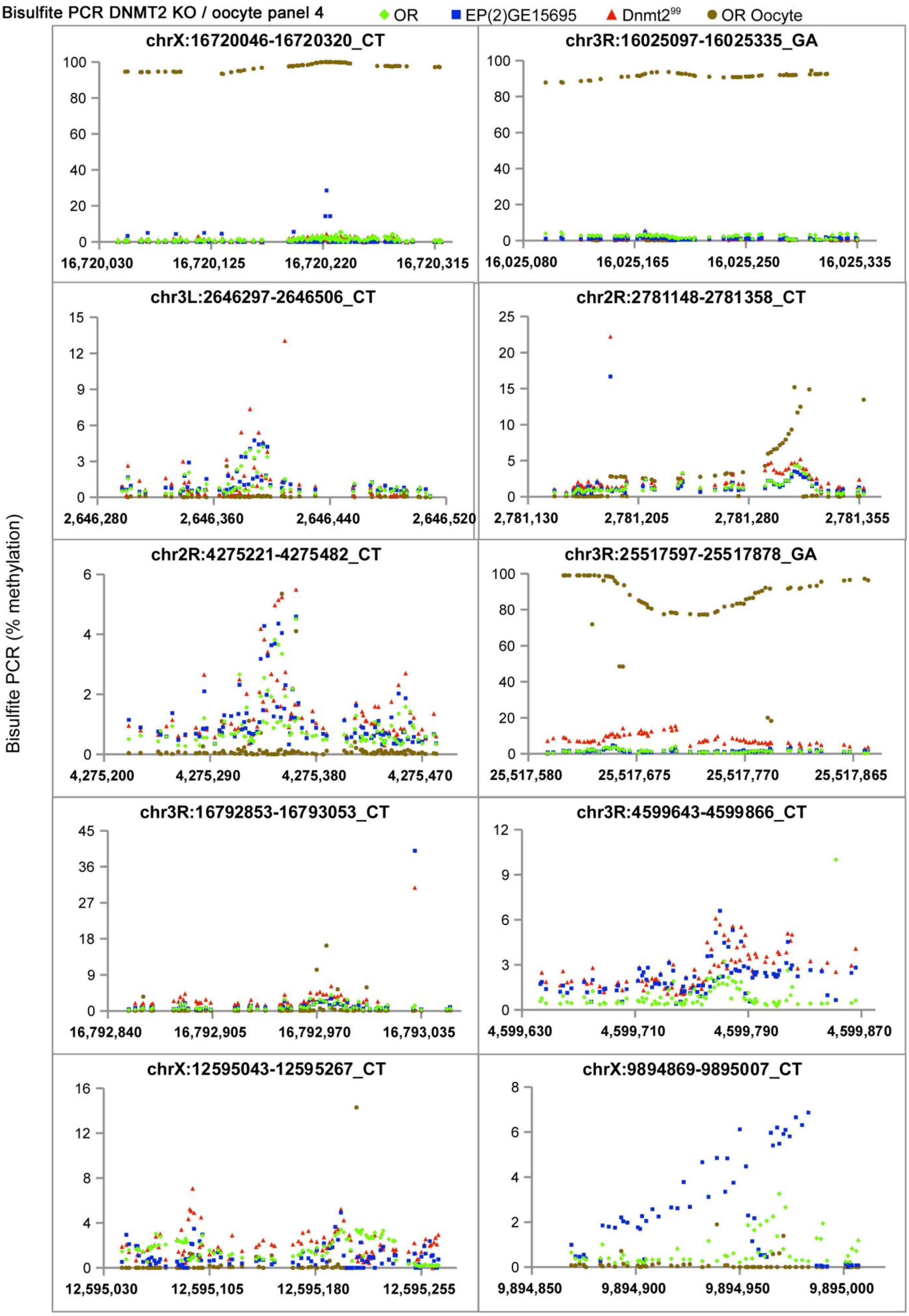
Bisulfite PCR DNMT2 KO / oocyte panel 1



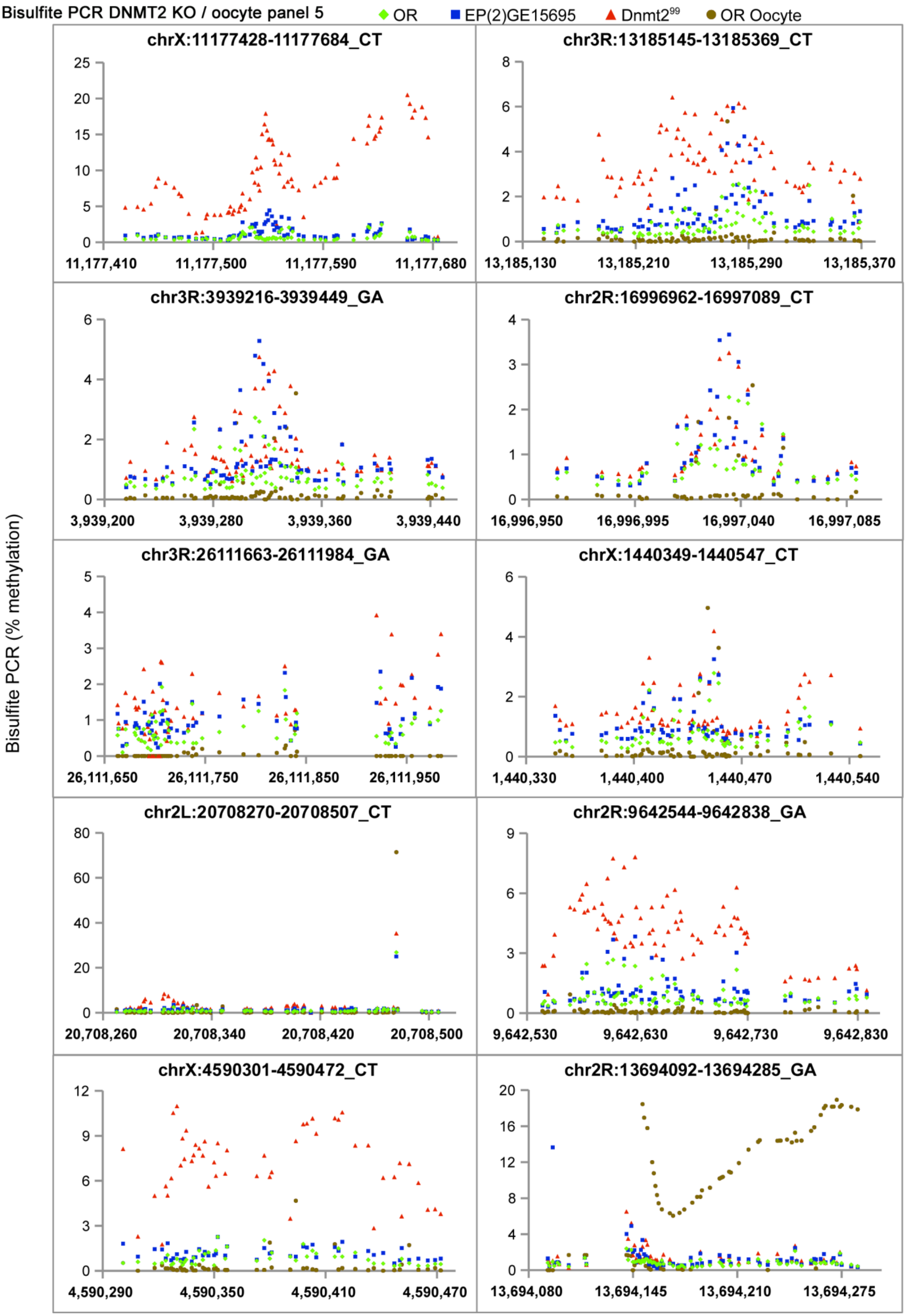
Bisulfite PCR DNMT2 KO / oocyte panel 2

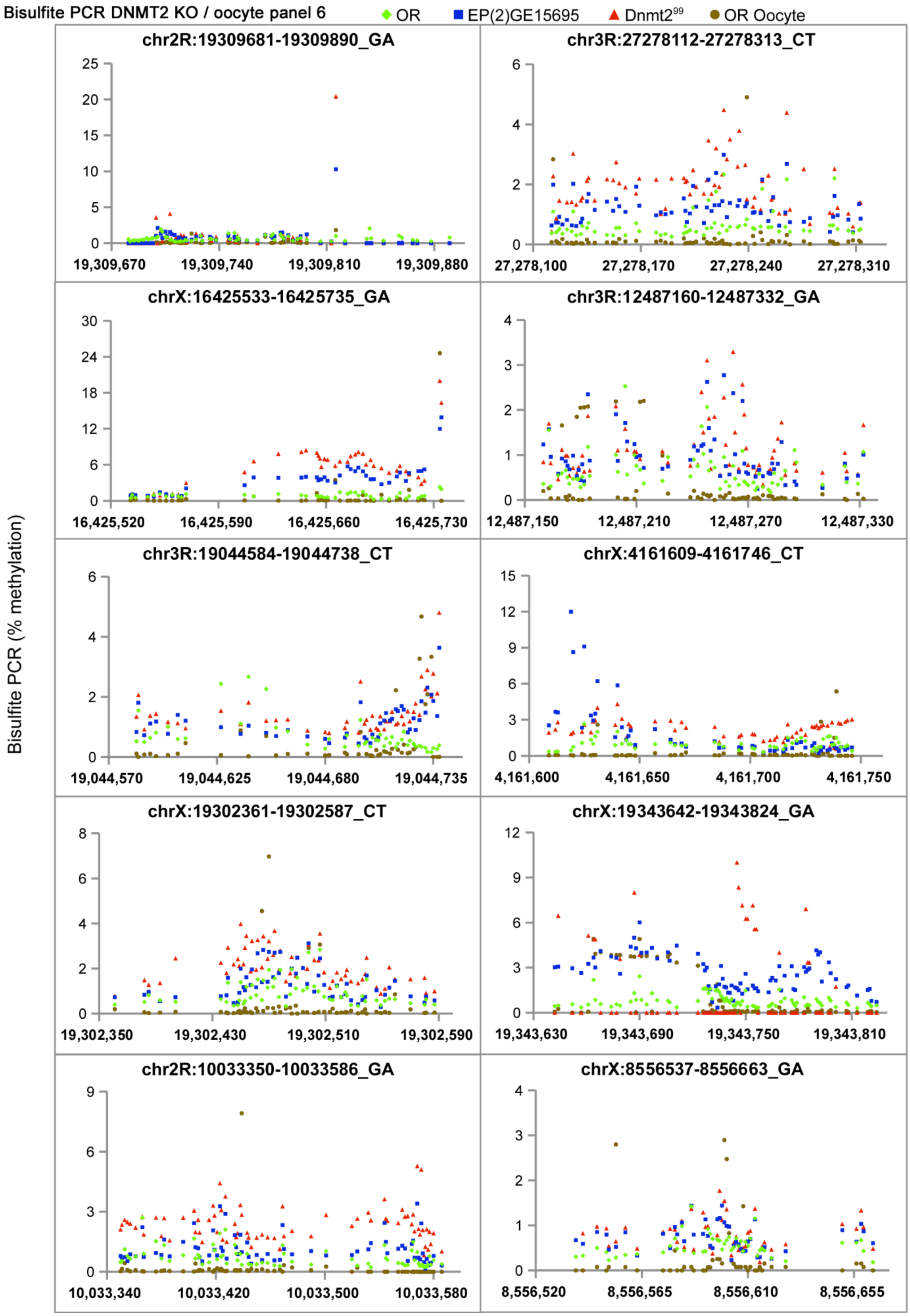



Bisulfite PCR DNMT2 KO / oocyte panel 3
◆ OR
■ EP(2)GE15695
▲ Dnmt2^99
● OR Oocyte
Bisulfite PCR (% methylation)
chrX:13468993-13469169_CT
0
3
6
9
12
15
13,468,980
13,469,040
13,469,100
13,469,160
chr2L:10124110-10124367_CT
0
1
2
3
4
5
10,124,100
10,124,190
10,124,280
10,124,370
chr2R:3112615-3112842_CT
0
2
4
6
8
10
3,112,600
3,112,680
3,112,760
3,112,840
chr3R:20916664-20916850_CT
0
4
8
12
16
20
20,916,650
20,916,715
20,916,780
20,916,845
chr3R:4274765-4274919_GA
0
2
4
6
8
4,274,750
4,274,805
4,274,860
4,274,915
chr3R:13571159-13571369_CT
0
3
6
9
12
13,571,150
13,571,220
13,571,290
13,571,360
chr2R:9642614-9642797_CT
0
1
2
3
4
9,642,600
9,642,660
9,642,720
9,642,780
chr3R:19898719-19898924_CT
0
4
8
12
16
20
19,898,710
19,898,780
19,898,850
19,898,920
chrX:16405496-16405737_CT
0
10
20
30
40
50
16,405,480
16,405,560
16,405,640
16,405,720
chrX:7045422-7045569_CT
0
20
40
60
80
100
7,045,410
7,045,465
7,045,520
7,045,575


Bisulfite PCR DNMT2 KO / oocyte panel 4
◆ OR
■ EP(2)GE15695
▲ Dnmt2^{99}
● OR Oocyte
Bisulfite PCR (% methylation)
chrX:16720046-16720320_CT
0 20 40 60 80 100
16,720,030 16,720,125 16,720,220 16,720,315
chr3R:16025097-16025335_GA
0 20 40 60 80 100
16,025,080 16,025,165 16,025,250 16,025,335
chr3L:2646297-2646506_CT
0 3 6 9 12 15
2,646,280 2,646,360 2,646,440 2,646,520
chr2R:2781148-2781358_CT
0 5 10 15 20 25
2,781,130 2,781,205 2,781,280 2,781,355
chr2R:4275221-4275482_CT
0 2 4 6
4,275,200 4,275,290 4,275,380 4,275,470
chr3R:25517597-25517878_GA
0 20 40 60 80 100
25,517,580 25,517,675 25,517,770 25,517,865
chr3R:16792853-16793053_CT
0 9 18 27 36 45
16,792,840 16,792,905 16,792,970 16,793,035
chr3R:4599643-4599866_CT
0 3 6 9 12
4,599,630 4,599,710 4,599,790 4,599,870
chrX:12595043-12595267_CT
0 4 8 12 16
12,595,030 12,595,105 12,595,180 12,595,255
chrX:9894869-9895007_CT
0 2 4 6 8
9,894,850 9,894,900 9,894,950 9,895,000

Bisulfite PCR DNMT2 KO / oocyte panel 5




Bisulfite PCR DNMT2 KO / oocyte panel 6
◆ OR
■ EP(2)GE15695
▲ Dnmt2^99
● OR Oocyte
Bisulfite PCR (% methylation)
chr2R:19309681-19309890_GA
0
5
10
15
20
25
19,309,670
19,309,740
19,309,810
19,309,880
chr3R:27278112-27278313_CT
0
2
4
6
27,278,100
27,278,170
27,278,240
27,278,310
chrX:16425533-16425735_GA
0
6
12
18
24
30
16,425,520
16,425,590
16,425,660
16,425,730
chr3R:12487160-12487332_GA
0
1
2
3
4
12,487,150
12,487,210
12,487,270
12,487,330
chr3R:19044584-19044738_CT
0
2
4
6
19,044,570
19,044,625
19,044,680
19,044,735
chrX:4161609-4161746_CT
0
3
6
9
12
15
4,161,600
4,161,650
4,161,700
4,161,750
chrX:19302361-19302587_CT
0
2
4
6
8
19,302,350
19,302,430
19,302,510
19,302,590
chrX:19343642-19343824_GA
0
3
6
9
12
19,343,630
19,343,690
19,343,750
19,343,810
chr2R:10033350-10033586_GA
0
3
6
9
10,033,340
10,033,420
10,033,500
10,033,580
chrX:8556537-8556663_GA
0
1
2
3
4
8,556,520
8,556,565
8,556,610
8,556,655

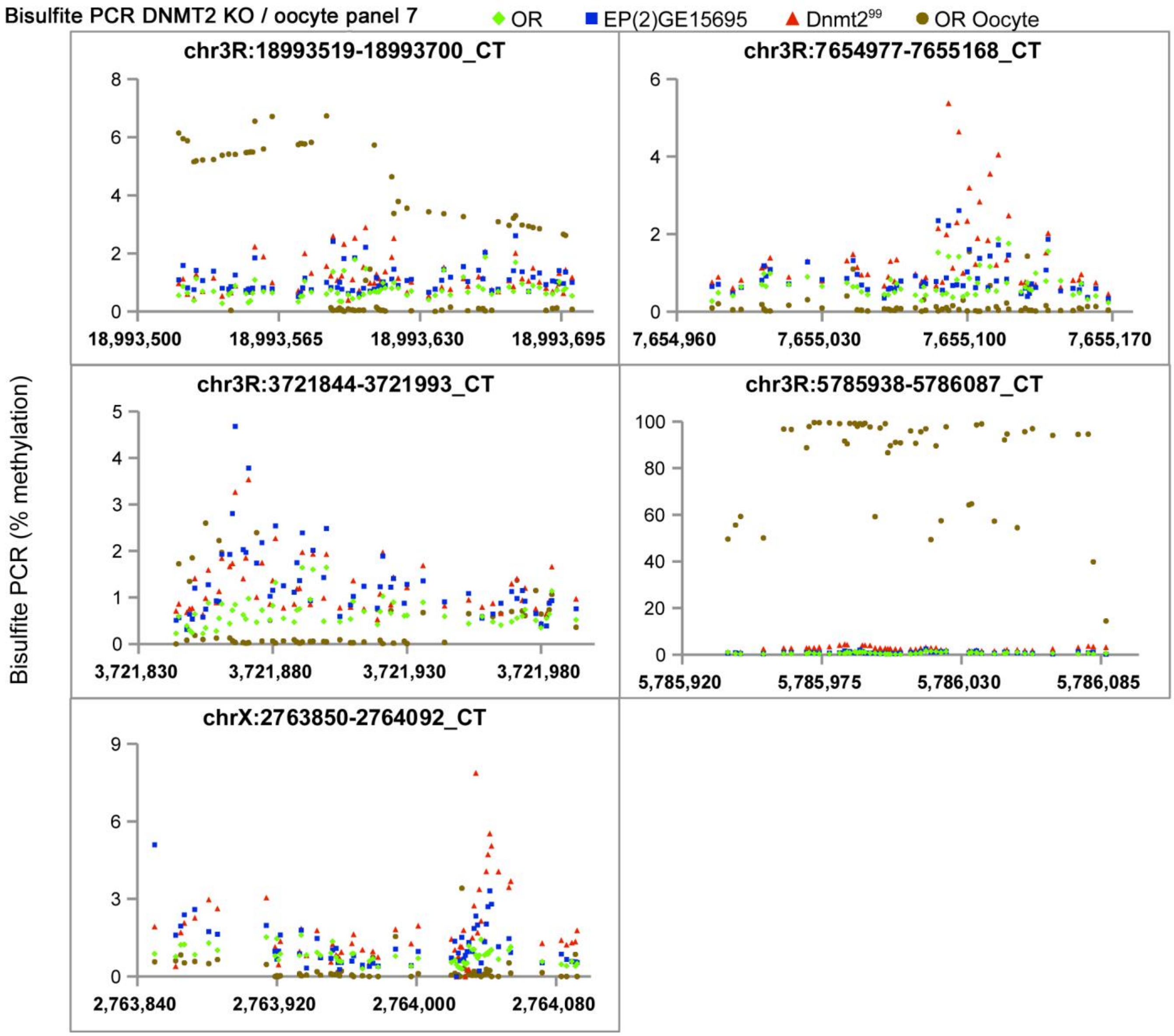


Figure S8. Methylation is present in flies deficient for the DNA methyltransferase MT2 and at some loci in unfertilized oocytes. The full set of 66 regions that were analyzed is shown. Methylated regions identified by MeDIP-bisulfite sequencing were PCR amplified from bisulfite converted DNA and Illumina sequenced to at least 10,000X coverage. Each dot represents one cytosine: green – bisulfite PCR (same data as in Figure S5); brown – unfertilized oocyte; red – *Mt2* deficient; blue – EP(2)GE15695 (*Mt2* wild type). The y-axis indicates the percent of methylated cytosines in the bisulfite PCR.

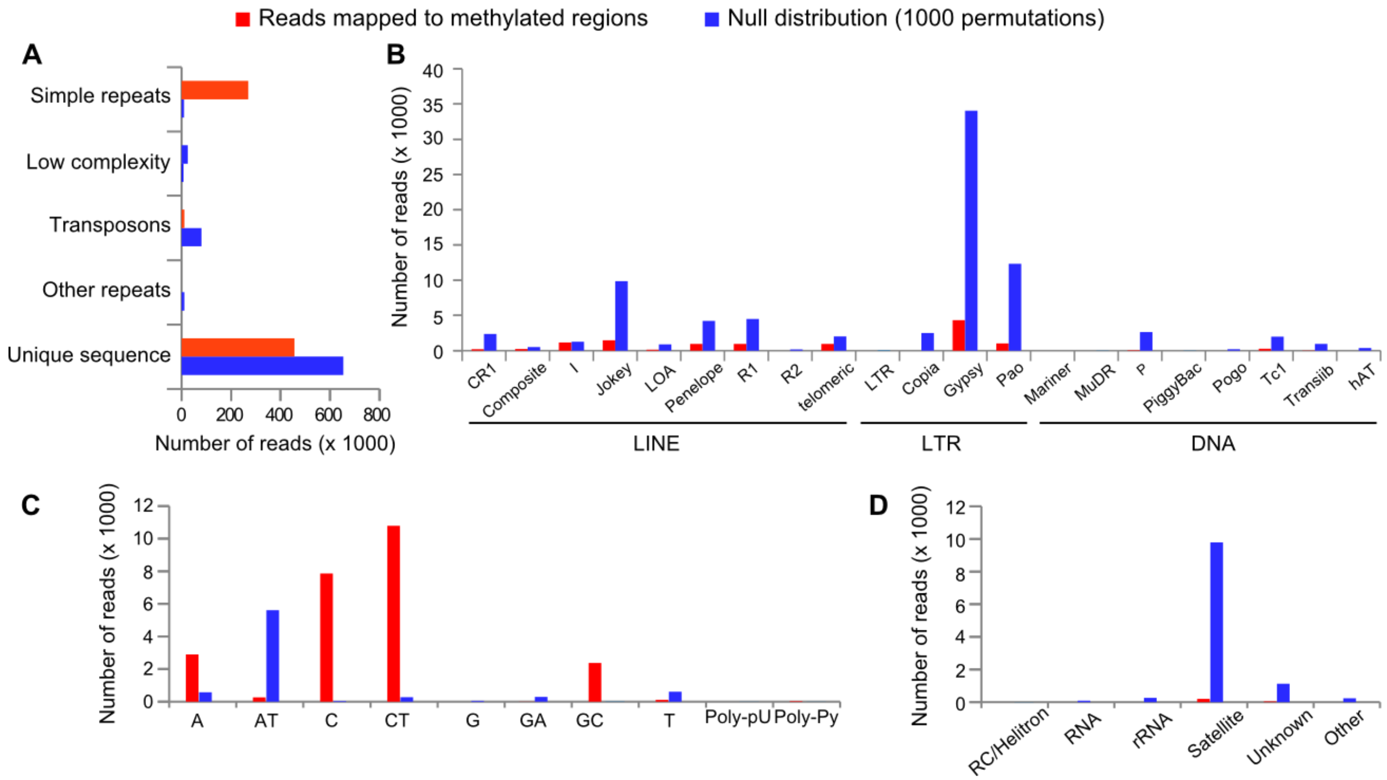


Figure S9. Sequence properties of methylated regions. The number of sequence reads within methylated regions that overlap with various sequence classes (red bars); blue bars represent the average and standard deviation of 1,000 randomized permutations of the same number of reads. **A.** Distribution of methylation between unique and repeat sequences. Methylation is much more likely to present in simple sequence repeats, and less likely to be present in transposons or unique sequences. **B-D.** Methylation of transposons and other repeat types. Methylation is depleted in all transposon families except the I element (B), enriched in some types of low-complexity sequence (C), and depleted from RNA, satellite, and other repeats (D). **Methods:** The repeat sequence annotation for the *D. melanogaster* dm3 assembly was downloaded from the UCSC Table Browser table:rmsk. The repeat annotation was intersected with the 762,655 primary alignments that align by at least 51% to the 25,497 methylated regions The intersection was obtained with intersectBed from the BEDTools suite (Quinlan and Hall 2010), run with -f0.51 option which requires that at least 51% of a read overlaps an annotated repeat. We used primary reads rather than methylated regions because of the difficulty in mapping a read to a specific repeat element. The results were compared to a random expectation distribution. We used shuffleBed from the BEDTools suite to randomly permute the locations of the 762,655 primary alignments. We used the –chrom option, which keeps the alignments on the same chromosome and only randomizes their location on the chromosome. A distribution of random annotations was generated by 1,000 repetitions of the permutation procedure, from which we calculated the mean and standard deviation.

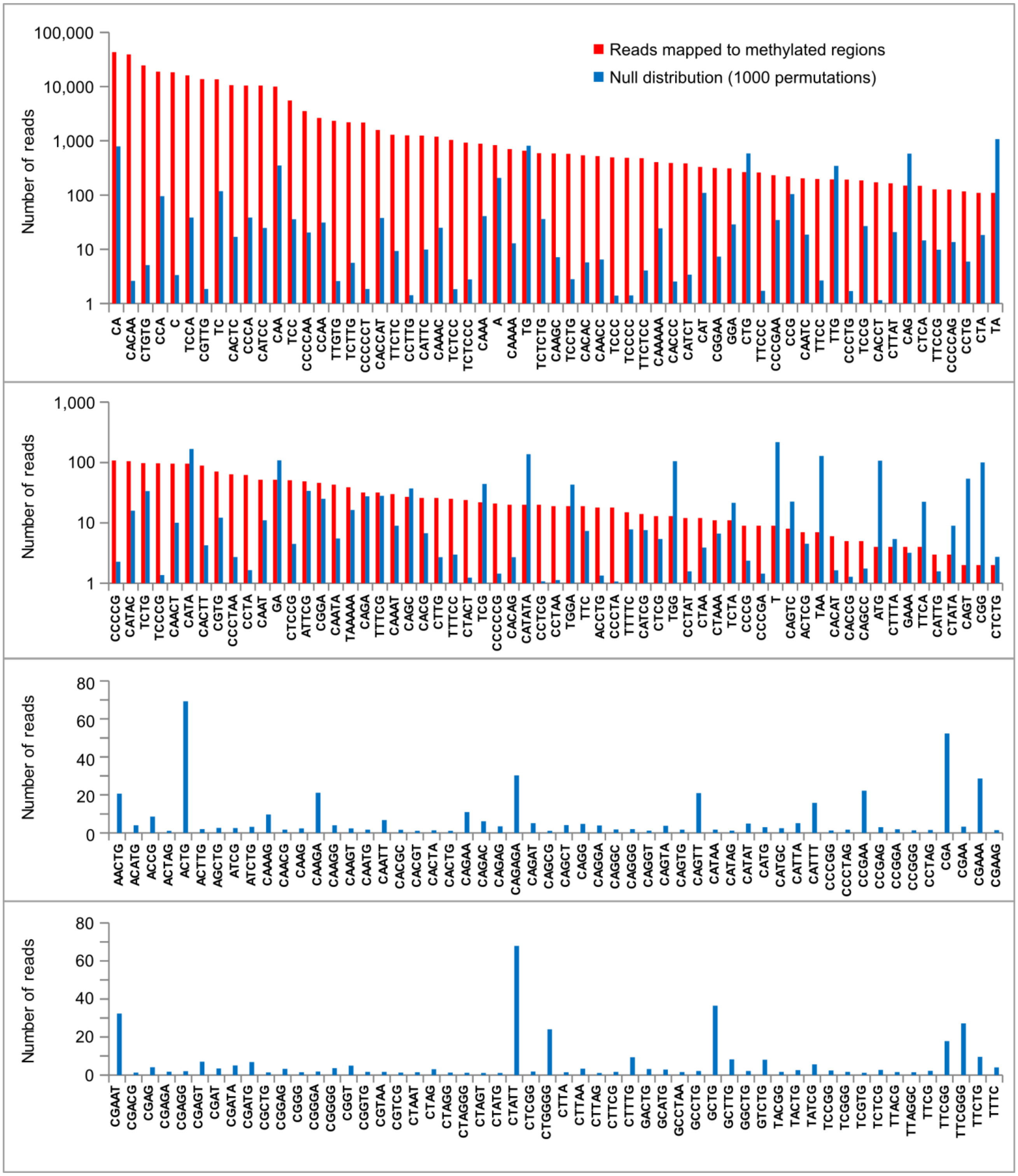


Figure S10. Methylation of various classes of simple sequence repeats. Red bars represent the number of sequence reads within methylated regions that overlap with a given simple sequence repeat; blue bars represent the average and standard deviation of 1,000 randomized permutations of the reads. Simple sequence repeats that lack a cytosine are not displayed. The scale on the y-axis of the top two panels is logarithmic. Few of the simple sequence repeats in which methylation is enriched contain Gs, but some of these are highly enriched.

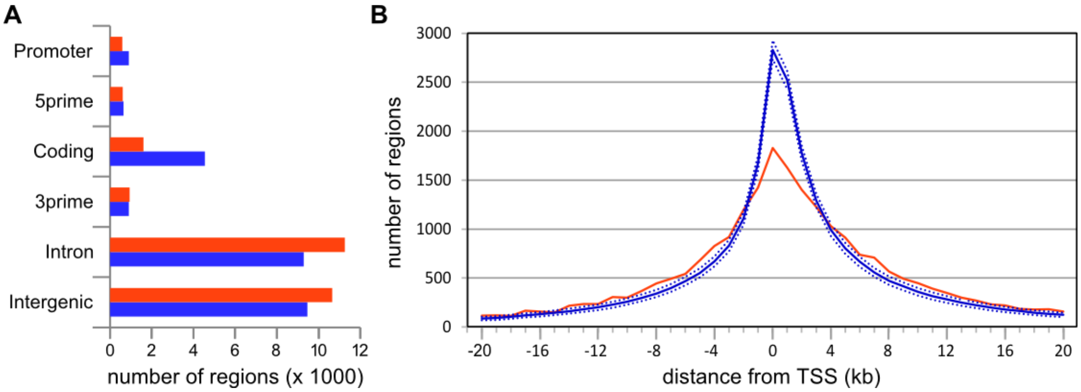


Figure S11. Methylated regions and genic features. A. The number of methylated regions overlapping with annotation features in the *Drosophila* genome (red bars). The blue bars represent the average and standard deviation of 1,000 randomized permutations of the 25, 319 methylated regions. Compared to the random selection of regions, methylated regions are more likely to be found in introns and intergenic regions, and less likely to be found at promoters and within coding regions. **B.** Distance of methylated regions from the nearest transcription start site. For each of the 25, 319 methylated regions, we calculated the distance to the nearest annotated transcription start site (TSS). The red line shows the number of methylated regions at a given distance from the nearest TSS. The solid blue line indicates the mean distance to the nearest TSS of 1,000 random permutations of the genomic locations of the 25, 319 methylated regions. The dotted blue lines denote the 95.6% confidence intervals. This analysis shows a depletion of methylated regions near TSSs. **Methods:** The gene annotation for the *D. melanogaster* dm3 assembly was downloaded in BED format from the UCSC Table Browser table:flyBaseGene. Non redundant files for the various gene annotation features (promoter, 5' UTR, coding exon, 3' UTR) were obtained by collapsing all features with overlapping coordinates; a promoter was defined as the sequence up to 300bp upstream of a transcription start sites. Regions that were annotated as more than one feature (e.g.: as 5' UTR and promoter) were retained independently. Introns were defined as the sequences within a gene that did not correspond to any exon. Intergenic regions were defined as the genome sequences that did not correspond to an intron or any other gene annotation feature. The degree of overlap between methylated regions and gene annotation features was determined with intersectBed from the BEDTools suite, run with -f0.51 option, which requires that at least 51% of a methylated region overlaps a gene annotation feature. The results were compared to a random expectation distribution. We used shuffleBed from the BEDTools suite to randomly permute the locations of the 25,497 methylated regions. We used the –chrom option, which keeps the regions on the same chromosome and only randomizes their location on the chromosome. The random expectation distribution was generated by 1,000 repetitions of the permutation procedure and by intersecting each repetition with the gene annotation features.

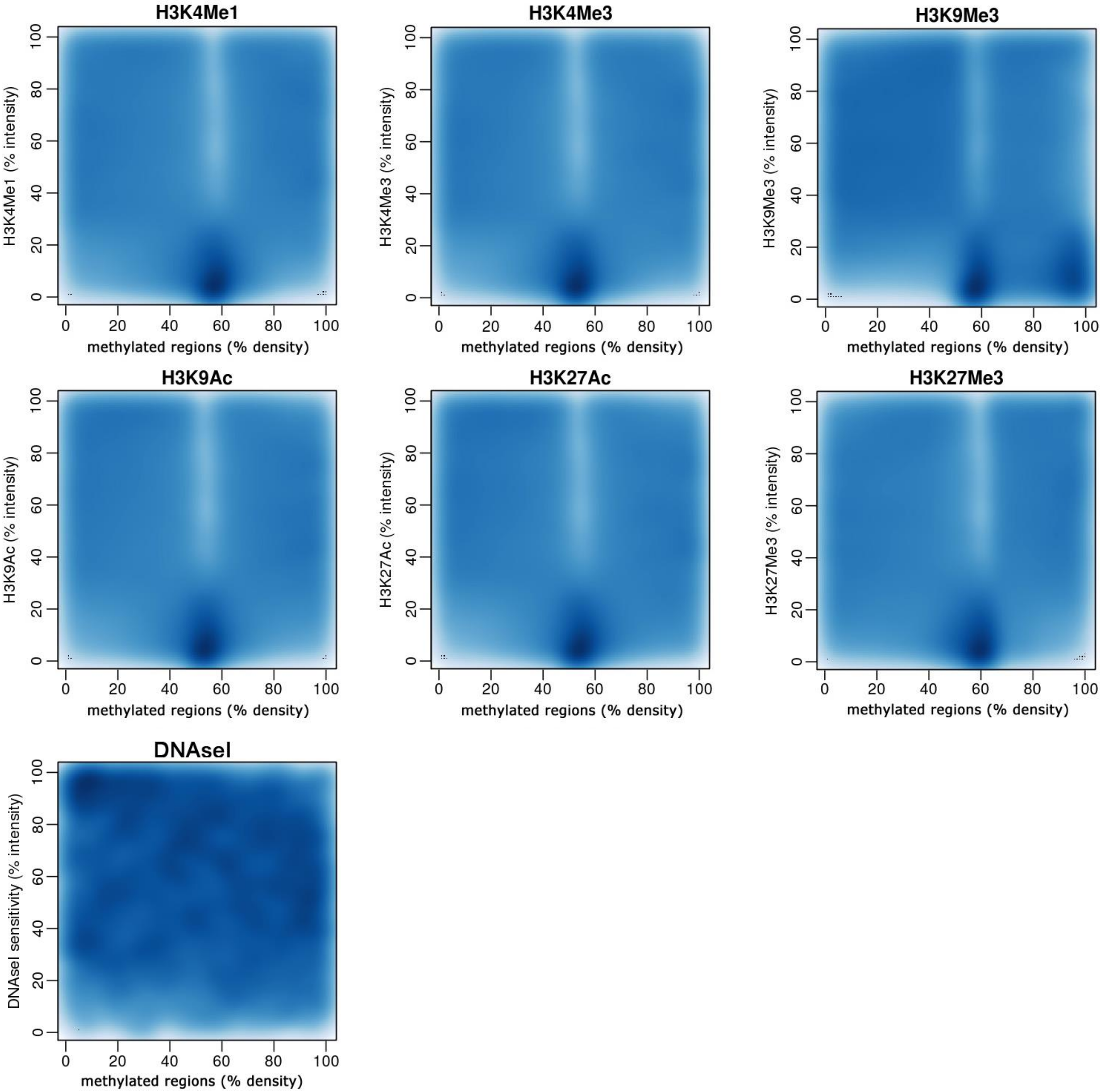


Figure S12. Lack of correlation between methylation and chromatin features. Scatterplots comparing the density of methylated region (x-axis) with the intensities of various histone tail modifications and of DNase I hypersensitivity (y-axis). The scatterplots illustrate a general lack of correlation. **Methods:** The density distribution of methylated regions over 100kb intervals was determined using fseq with the -l100000 -s100 options. Histone tail modification data for *D. melanogaster* (developmental stage: E0-4h) were downloaded as wiggle files from the GEO database with the accession numbers: GSM400656 (H3K4Me3), GSM401407 (H3K27Ac), GSM401408 (H3K9Ac), GSM401409 (H3K4Me1), GSM439448 (H3K27Me3), GSM439457 (H3K9Me3). DNase I sensitivity data for *D. melanogaster* (developmental stage 5) was downloaded in BED format from the UCSC Table Browser table:bdtnpDnaseAccS5, dm3 assembly, and converted to a density distribution using fseq with the -l100000 -s100 options. The intensities of the distributions of methylated regions and chromatin state features were percentile-normalized (histone tail modification data with a value of '0' were skipped during normalization) and compared with the normalized distribution of methylation density by scatterplot using the 'smoothscatter' package in R.