Supplemental material

Figure S1: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing at least one repeat 1000 bp upstream and downstream of the retrocopy.
Figure S2: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing no repeat 1000 bp upstream and downstream of the retrocopy.
Figure S3: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing at least one repeat 1000 bp downstream of the retrocopy.
Figure S4: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing at least one repeat 1000 bp upstream of the retrocopy.
Figure S5-S21: IGV screenshots of 17 differentially methylated CGIs (methylome 1).
Figure S22: Expression analysis of IGF2BP3 in placenta of different trimester of pregnancy

Table S1: Classification of CGIs overlapping retrogenes
Table S2: Expression level of retrogenes overlapping a CGI as classified by the UCSC browser
Table S3: Genomic location of retrogenes overlapping a CGI according to RefSeq genes
Table S4: Genomic location of retrogenes overlapping a CGI according to RefSeq genes according to their expression level
Table S5: Expression level of intergenic retrogenes overlapping a CGI according to their methylation level
Table S6: Expression level of intragenic retrogenes overlapping a CGI according to their methylation level
Table S7: CGI methylation levels in other tissues
Table S8: Primer sequences for genomic sequence analysis
Table S9: Primer sequences for deep bisulfite sequencing (universal tags in bold)
Table S10: Primer sequences for RNA analysis
Table S11: Detailed description of the analyzed 17 RO-CGIs
Table S12: Row data of SNaPshot analysis

Interpretation of Phred Quality Score

Strategy of genomic primer design

Strategy of bisulfit primer design

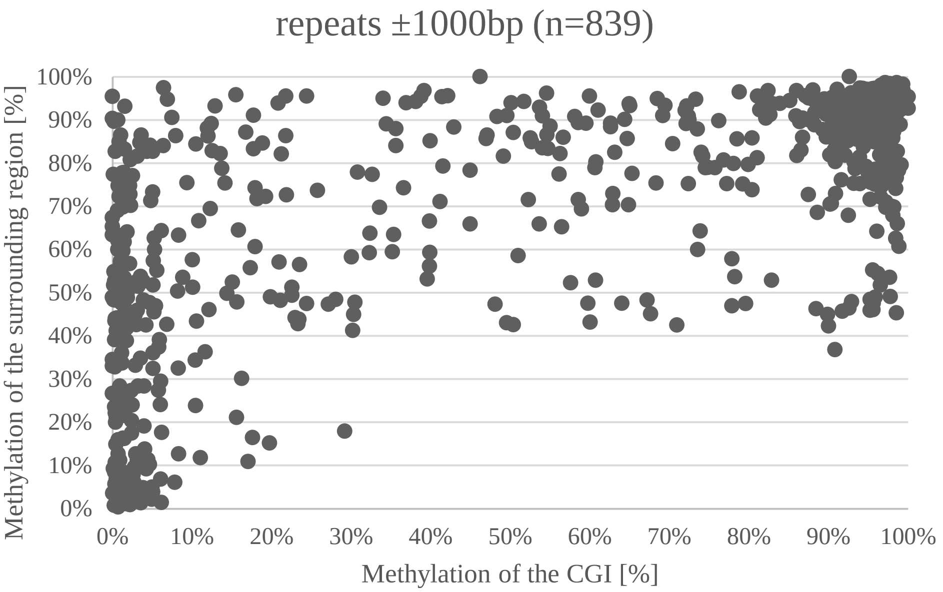


Fig. S1: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing at least one repeat 1000 bp upstream and downstream of the retrocopy.

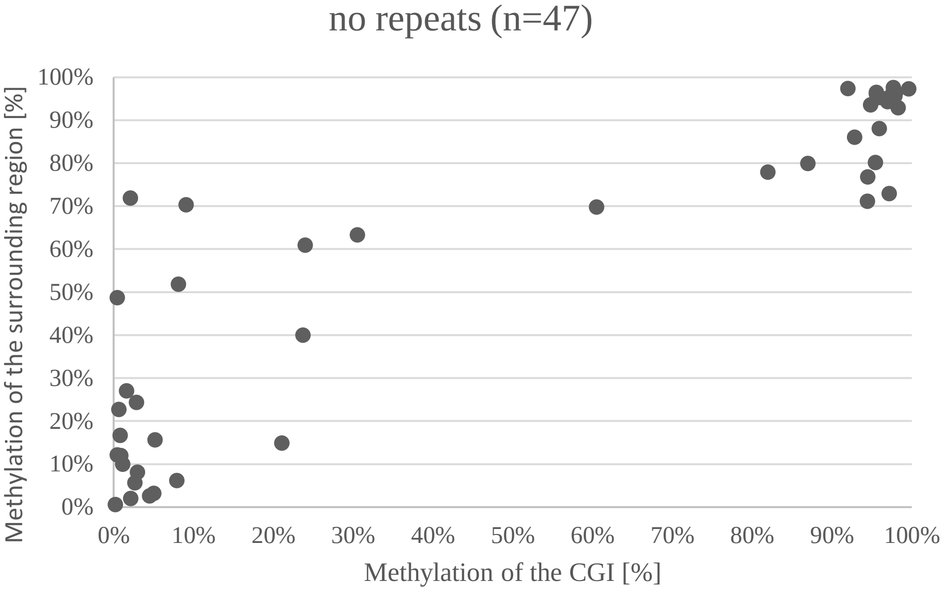
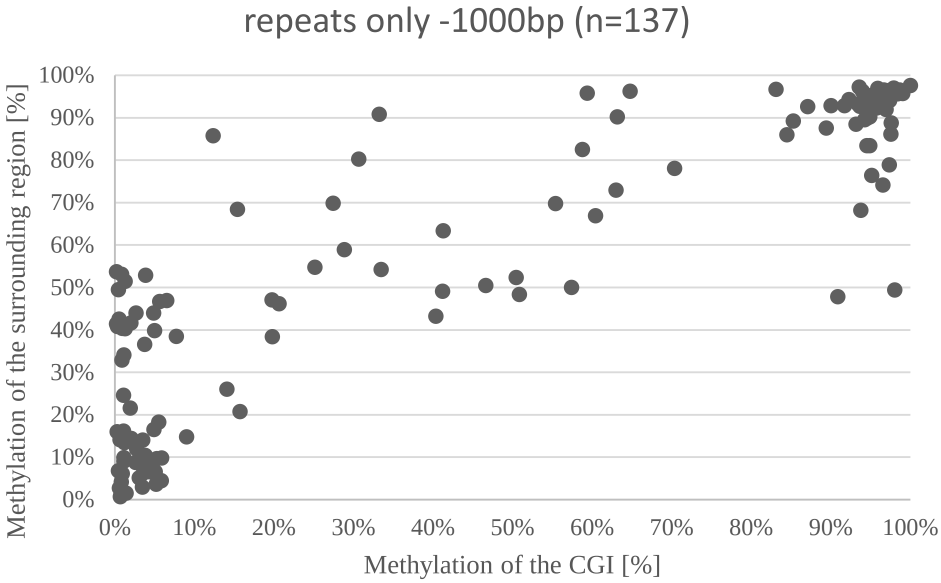
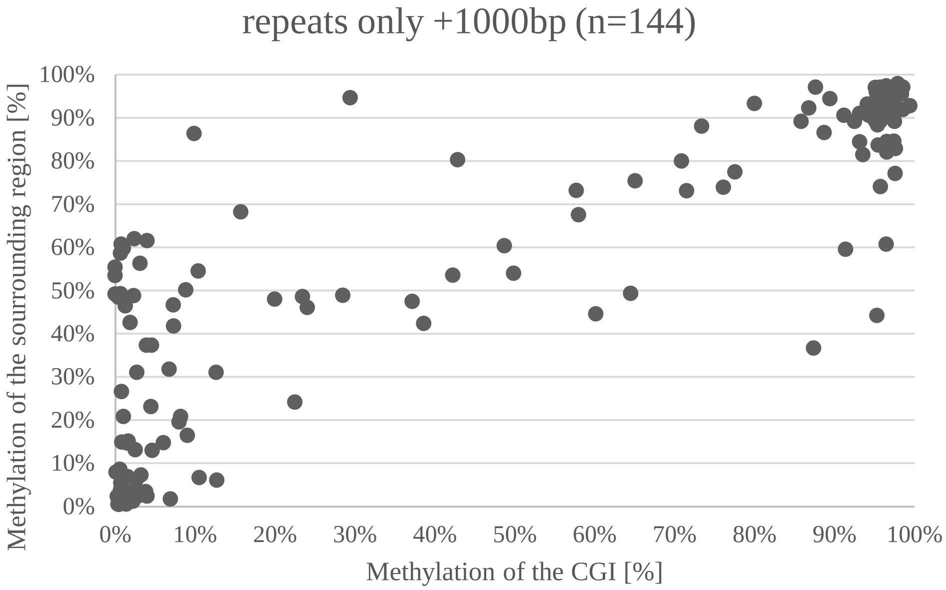


Fig. S2: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing no repeat 1000 bp upstream and downstream of the retrocopy.



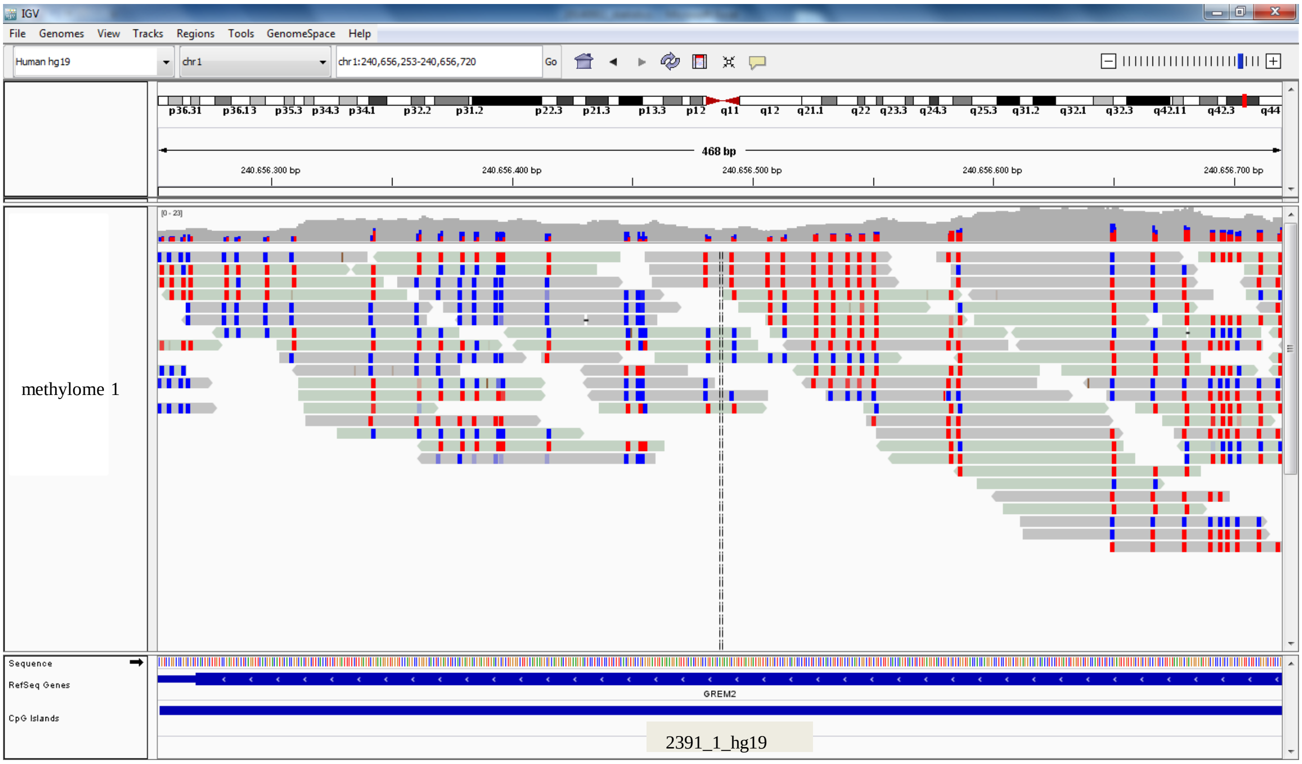
repeats only -1000bp (n=137)

Fig. S3: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing at least one repeat 1000 bp downstream of the retrocopy.



repeats only +1000bp (n=144)

Fig. S4: Methylation level of the CGI overlapping a retrocopy plotted against the methylation level of the surrounding region containing at least one repeat 1000 bp upstream of the retrocopy.

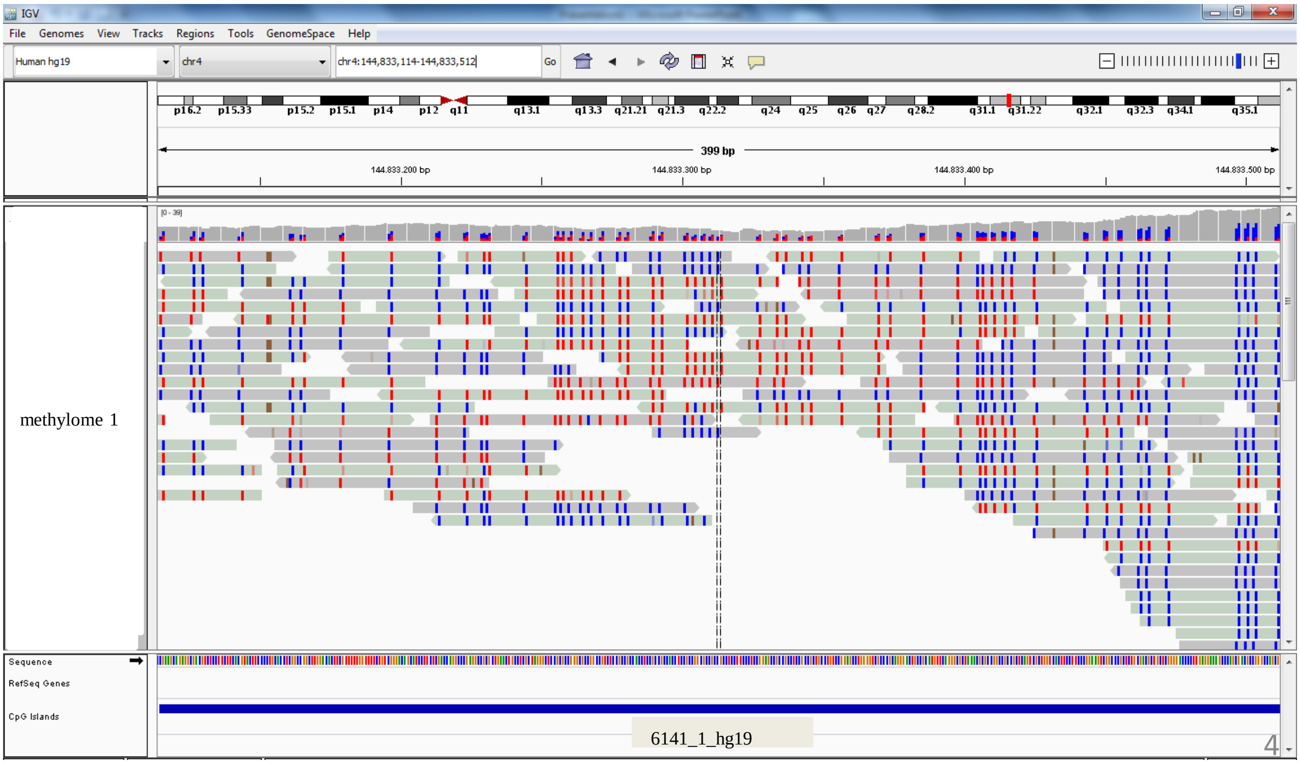


Human hg19

chr1

chr1:240,656,253-240,656,720

Fig. S5: Screenshot IGV Browser of CGI 2391_1_hg19 (*GREM2*)

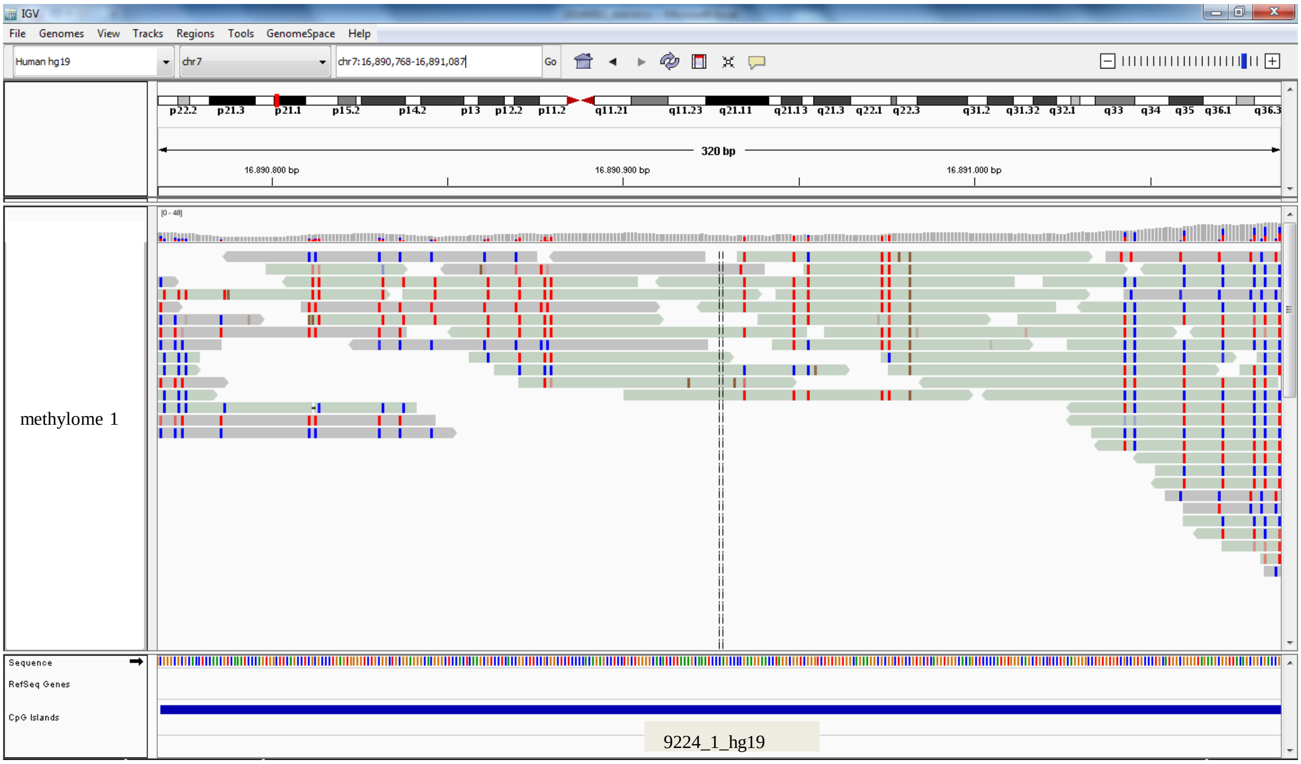


Human hg19

chr4

chr4:144,833,114-144,833,512

Fig. S6: Screenshot IGV Browser of CGI 6141_1_hg19 (*LOC101927636*)

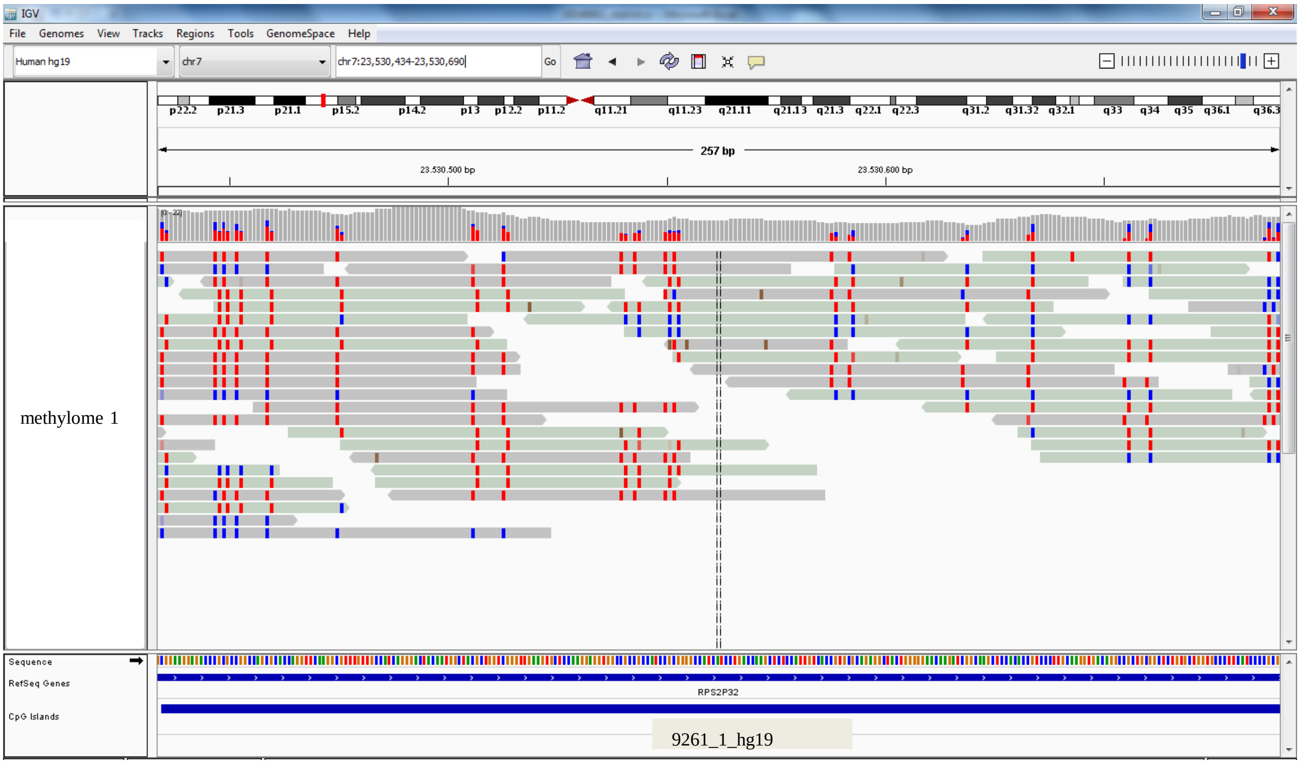


Human hg19

chr7

chr7:16,890,768-16,891,087

Fig. S7: Screenshot IGV Browser of CGI 9224_1_hg19 (intergenic)



Human hg19

chr7

chr7:23,530,434-23,530,690

Fig. S8: Screenshot IGV Browser of CGI 9261_1_hg19 (*RPS2P32*)

Fig. S9: Screenshot IGV Browser of CGI 9377_1_hg19 (intergenic)

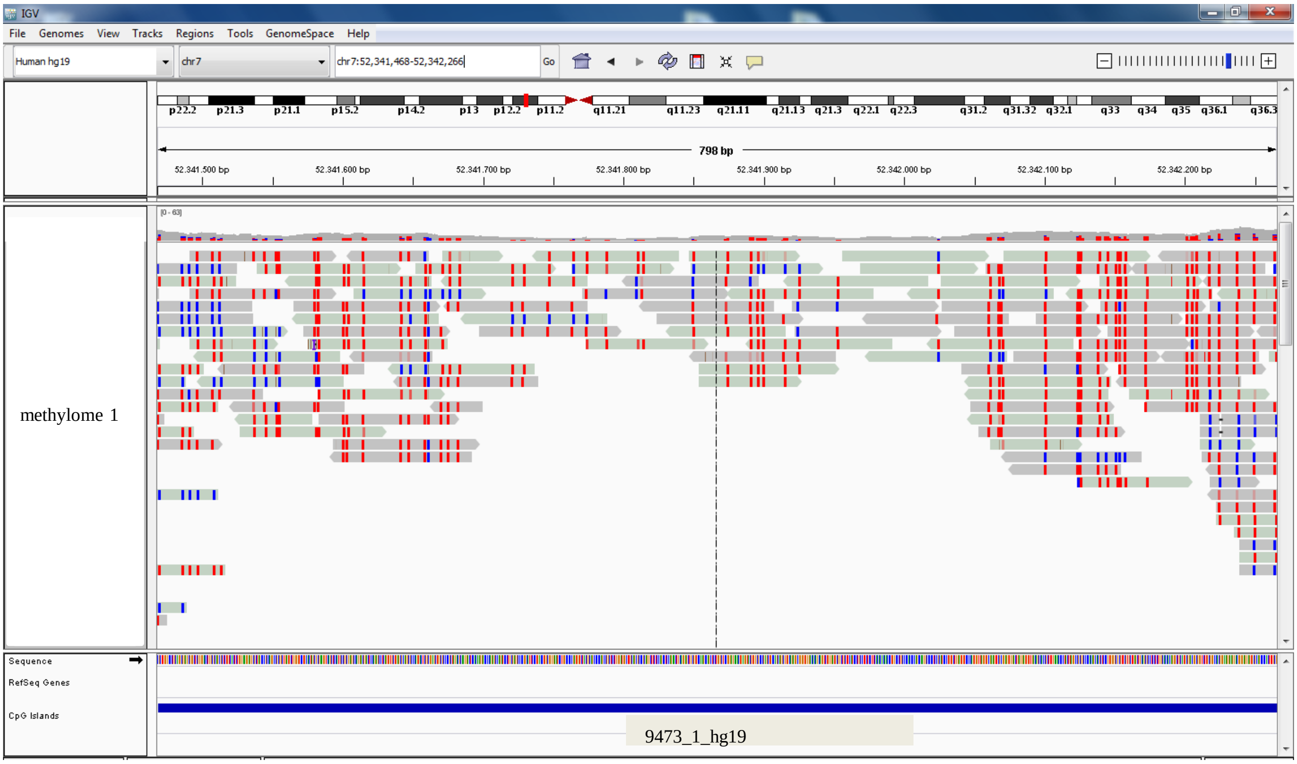
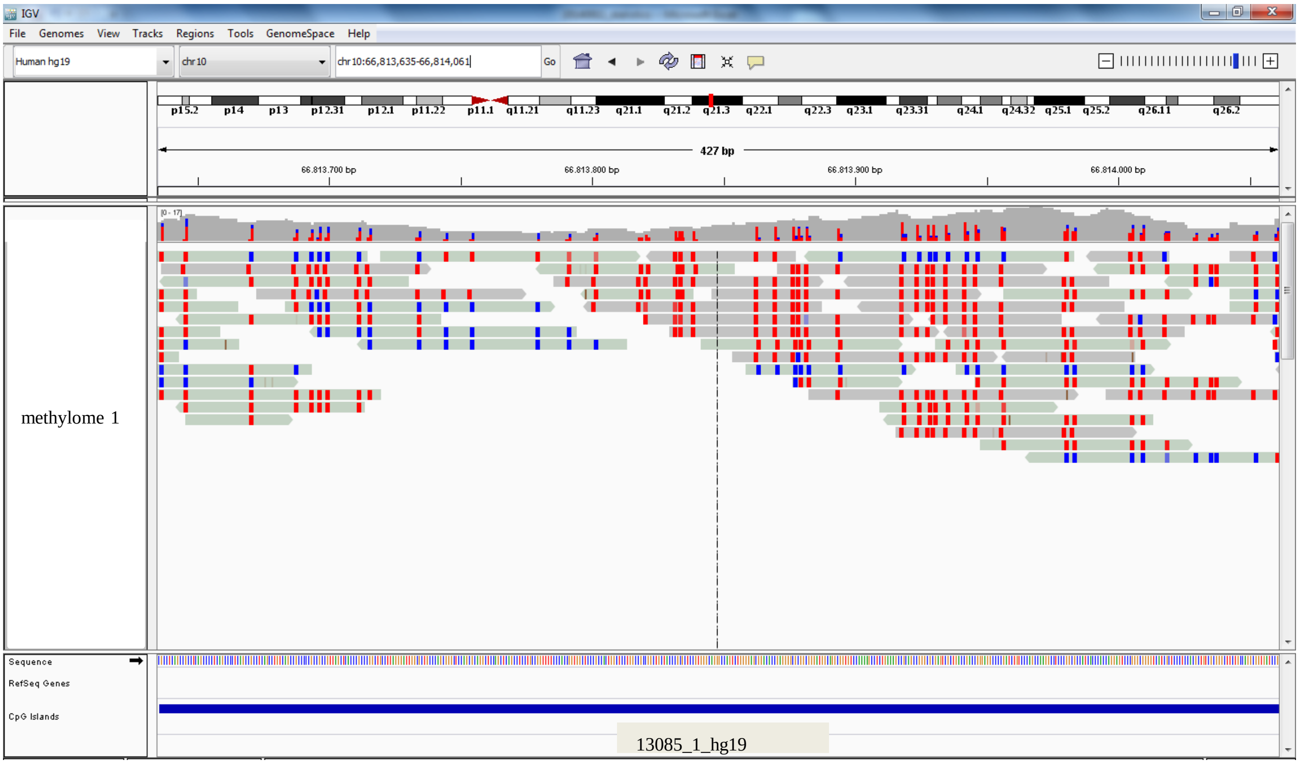


Fig. S10: Screenshot IGV Browser of CGI 9473_1_hg19 (intergenic)



Human hg19

chr10

chr10:66,813,635-66,814,061

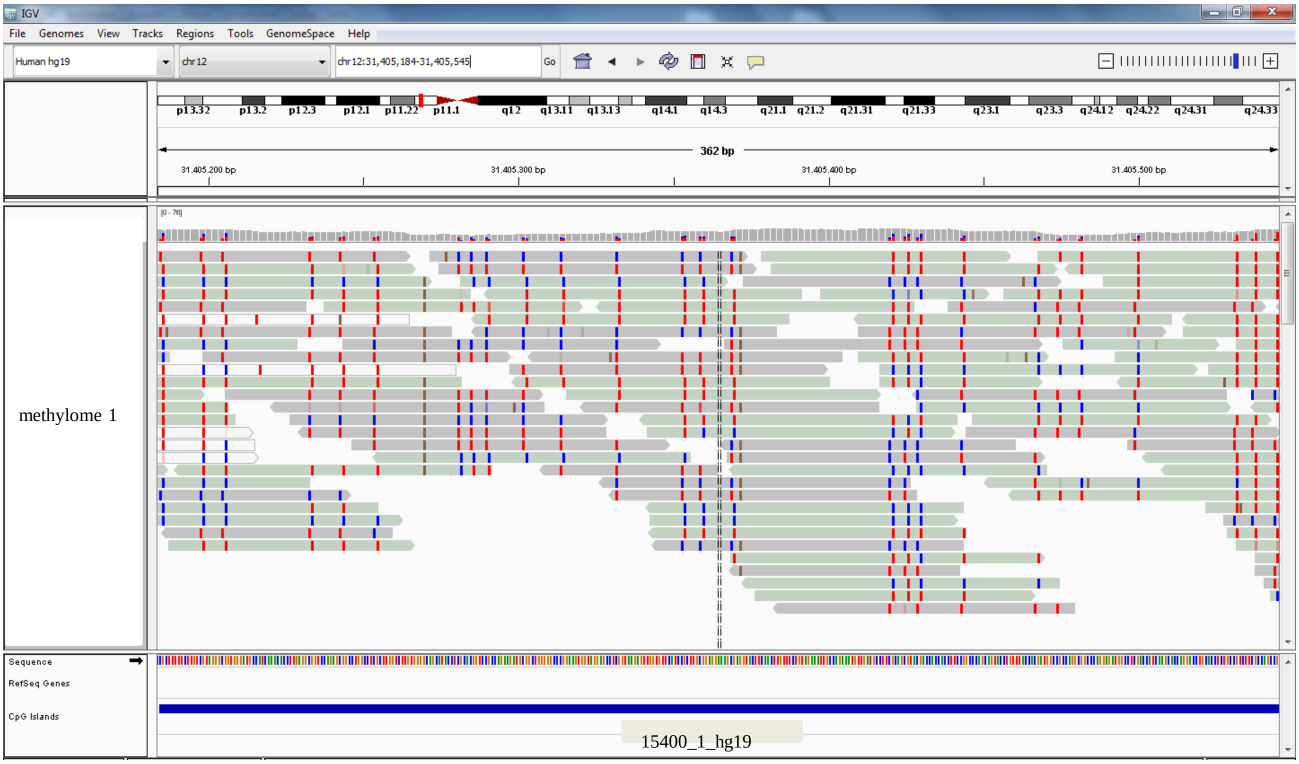
Fig. S11: Screenshot IGV Browser of CGI 13085_1_hg19 (intergenic)

Human hg19

chr10

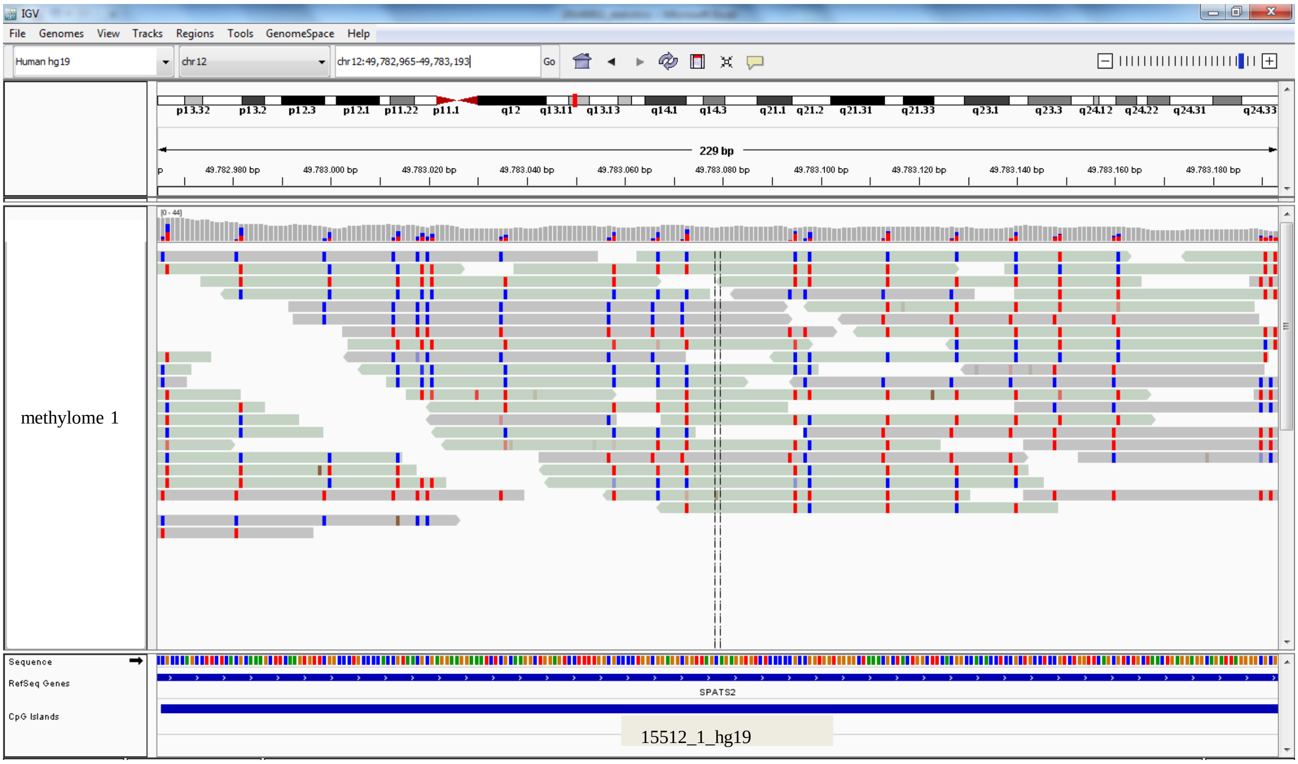
chr10:91,596,974-91,597,792

Fig. S12: Screenshot IGV Browser of CGI 13250_1_hg19 (*LINC00865*)



Human hg19
chr12
chr12:31,405,184-31,405,545

Fig. S13: Screenshot IGV Browser of CGI 15400_1_hg19 (intergenic)



Human hg19
chr12
chr12:49,782,965-49,783,193

Fig. S14: Screenshot IGV Browser of CGI 15512_1_hg19 (*SPATS2*)

Human hg19

chr13

chr13:21,893,156-21,893,605

Fig. S15: Screenshot IGV Browser of CGI 16448_1_hg19 (*MIPEPP3*)

Human hg19

chr13

chr13:23,412,207-23,412,623

Fig. S16: Screenshot IGV Browser of CGI 16458_1_hg19 (intergenic)

Fig. S17: Screenshot IGV Browser of CGI 17031_1_hg19 (intergenic)

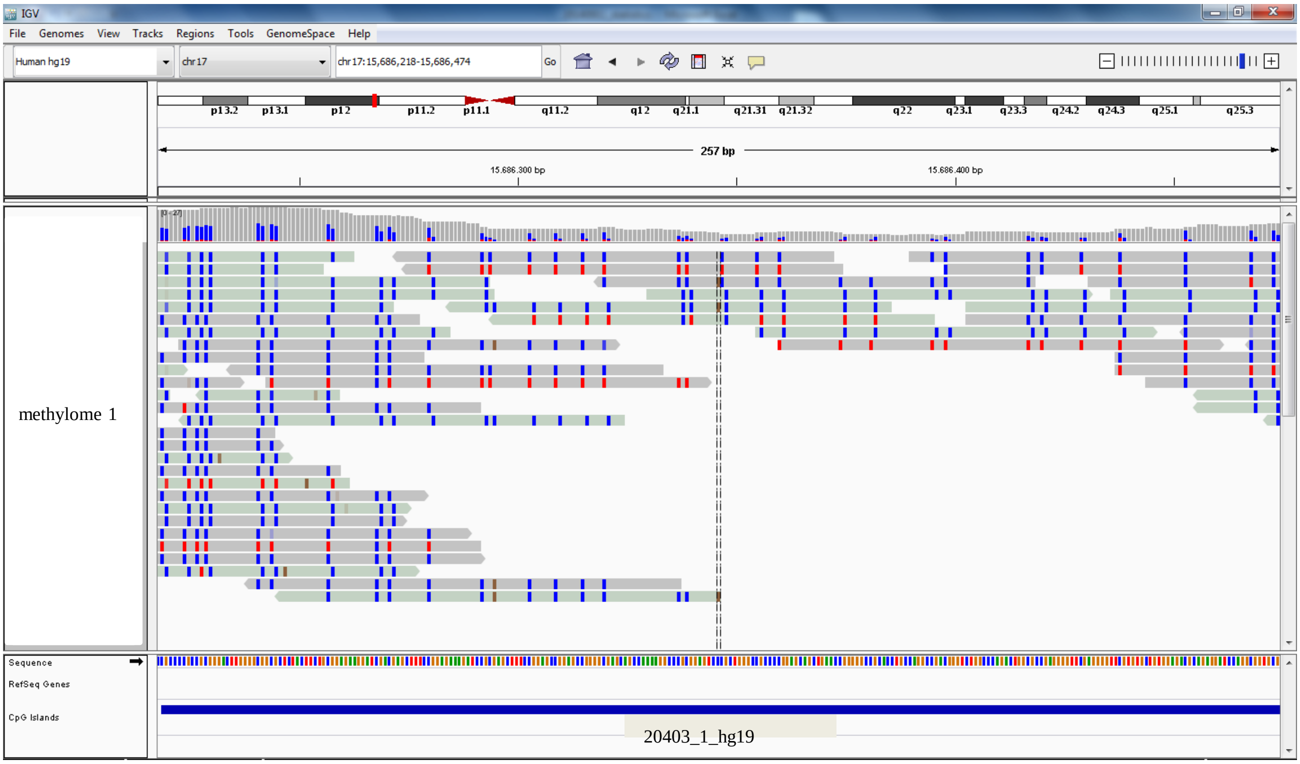


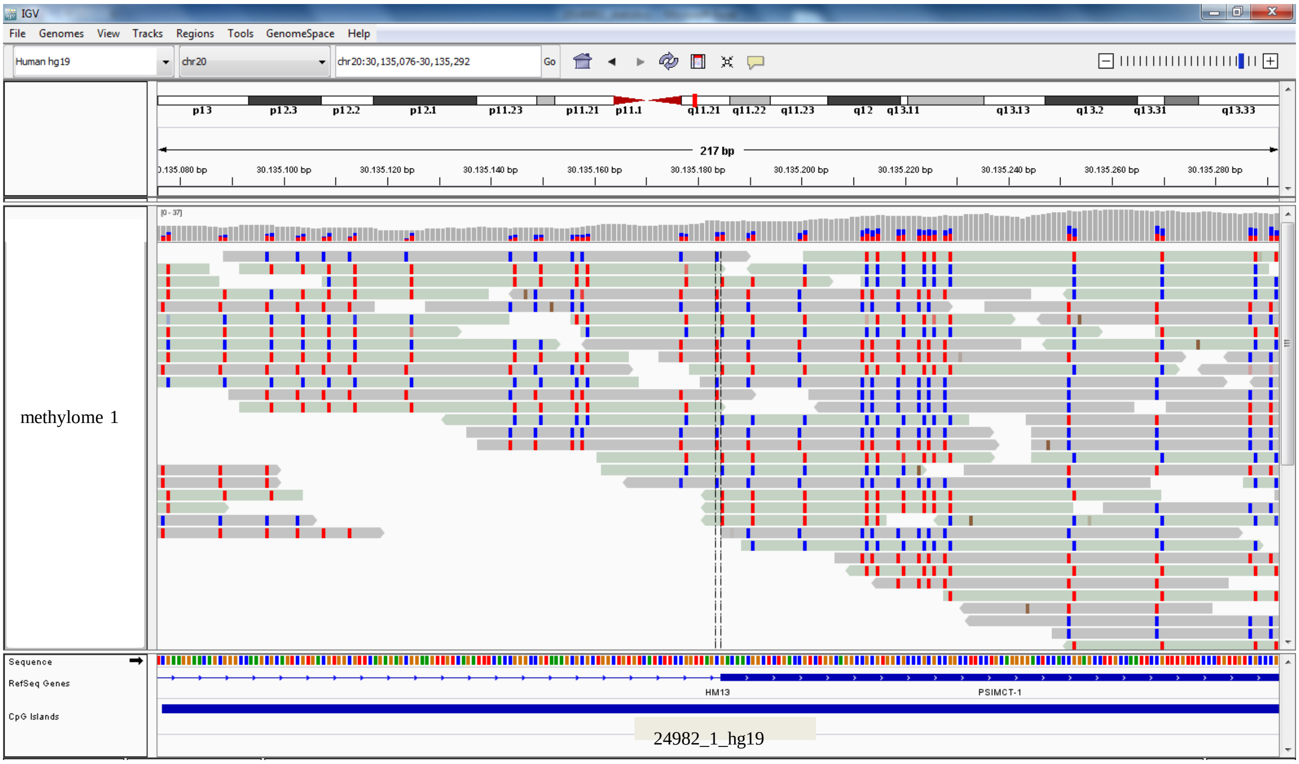
Fig. S18: Screenshot IGV Browser of CGI 20403_1_hg19 (intergenic)


Human hg19

chr19

chr19:21,860,792-21,861,016

Fig. S19: Screenshot IGV Browser of CGI 23548_1_hg19 (intergenic)


Human hg19

chr20

chr20:30,135,076-30,135,292

Fig. S20: Screenshot IGV Browser of CGI 24982_1_hg19 (*HM13*)

Human hg19

chrX

chrX:37,026,348-37,026,706

Fig. S21: Screenshot IGV Browser of CGI 26859_1_hg19 (*FAM47C*)

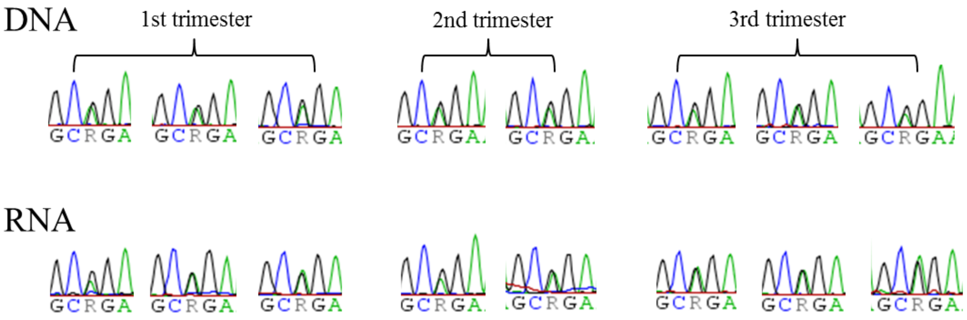


Fig. S22: Expression analysis of IGF2BP3 in placenta of different trimester. Both alleles are present in RNA, which indicates biallelic expression.

Table S1: Classification of CGIs overlapping retrogenes

number of occurences of one CGI/retrogene	number of ancestral genes	number of CGIs
1	541	541
2	134	268
3	41	123
4	26	104
5	8	40
6	7	42
7	7	49
8	2	16
9	4	36
10	1	10
11	2	22
12	0	0
13	3	39
14	0	0
15	0	0
16	1	16
not classifiable	n/a	13
sum	**778**	**1319**

Table S2: Expression level of retrogenes overlapping a CGI as classified by the UCSC browser

	expression level of retrogenes overlapping a CGI							
	expressed		**expressed noOrf**		**pseudogene**		**total number**	
number of retrogenes	934		188		197		1319	
methylation CGIs	[number]	[%]	[number]	[%]	[number]	[%]	[number]	[%]
CGIs m < 20%	273	29%	84	45%	27	14%	384	29%
CGIs 20% ≥ m ≤ 80%	125	13%	30	16%	40	20%	195	15%
CGIs m > 80%	436	47%	45	24%	107	54%	588	45%
not covered*	100	11%	29	15%	23	12%	152	12%

Table S3: Genomic location of retrogenes overlapping a CGI according to RefSeq genes

	location of the retrogenes overlapping a CGI					
	intergenenic (RefSeq)		**intragenic (RefSeq)**		**total**	
number of retrogenes	660		659		1319	
methylation CGIs	[number]	[%]	[number]	[%]	[number]	[%]
CGIs m < 20%	119	18%	265	40%	384	29%
CGIs 20% ≥ m ≤ 80%	124	19%	68	10%	192	15%
CGIs m > 80%	313	47%	278	42%	591	45%
not covered*	104	16%	48	7%	152	12%

Table S4: Genomic location of retrogenes overlapping a CGI according to RefSeq genes according to their expression level

	location of retrogenes overlapping a CGI					
	intergenenic (RefSeq)		intragenic (RefSeq)		total	
number of retrogenes	660		659		1319	
expression type	[number]	[%]	[number]	[%]	[number]	[%]
all 4 types expressed	**395**	**60%**	**539**	**82%**	**934**	**71%**
expressed	8	1%	10	2%	18	1%
expressed weak	327	50%	109	17%	436	33%
expressed strong	42	**6%**	119	18%	161	12%
expressed shuffle	18	3%	301	46%	319	24%
expressed noOrf	**90**	**14%**	**98**	**15%**	**188**	**14%**
pseudogene	**175**	**27%**	**22**	**3%**	**197**	**15%**

Table S5: Expression level of intergenic retrogenes overlapping a CGI according to their methylation level

	expression level of intergenic retrogenes overlapping a CGI							
	expressed		expressed shuffle		expressed strong		expressed strong noOrf	
number of retrogenes	8		18		42		14	
methylation CGIs	[number]	[%]	[number]	[%]	[number]	[%]	[number]	[%]
CGIs m < 20%	2	25%	3	17%	9	21%	7	50%
CGIs 20% ≥ m ≤ 80%	0	0%	2	11%	9	21%	1	7%
CGIs m > 80%	4	50%	8	44%	21	50%	1	7%
not covered*	2	25%	5	28%	3	7%	5	36%

	expression level of intergenic retrogenes overlapping a CGI							
	expressed weak		expressed weak noOrf		pseudogene		total number	
number of retrogenes	327		76		175		660	
methylation CGIs	[number]	[%]	[number]	[%]	[number]	[%]	[number]	[%]
CGIs m < 20%	49	15%	26	34%	22	13%	118	18%
CGIs 20% ≥ m ≤ 80%	65	20%	11	14%	38	22%	126	19%
CGIs m > 80%	162	50%	23	30%	93	53%	312	47%
not covered*	51	16%	16	21%	22	13%	104	16%

Table S6: Expression level of intragenic retrogenes overlapping a CGI according to their methylation level

	expression level of intragenic retrogenes overlapping a CGI							
	expressed		expressed shuffle		expressed strong		expressed strong noOrf	
number of retrogenes	10		301		119		48	
methylation CGIs	[number]	[%]	[number]	[%]	[number]	[%]	[number]	[%]
CGIs m < 20%	2	20%	136	45%	57	48%	21	44%
CGIs 20% ≥ m ≤ 80%	1	10%	23	8%	12	10%	15	31%
CGIs m > 80%	2	20%	122	41%	45	38%	11	23%
not covered*	5	50%	20	7%	5	4%	1	2%

	expression level of intragenic retrogenes overlapping a CGI							
	expressed weak		expressed weak noOrf		pseudogene		total number	
number of retrogenes	109		50		22		659	
methylation CGIs	[number]	[%]	[number]	[%]	[number]	[%]	[number]	[%]
CGIs m < 20%	14	13%	30	60%	5	23%	265	40%
CGIs 20% ≥ m ≤ 80%	12	11%	3	6%	2	9%	68	10%
CGIs m > 80%	74	68%	10	20%	14	64%	278	42%
not covered*	9	8%	7	14%	1	5%	48	7%

Table S7: CGI methylation levels in other tissues. For each CGI methylation (meth.) and coverage (cov.) are specified. **red:** methylation > 80%, **blue:** methylation < 20%, ***** data published in Rademacher et al. (2014), ****** data published by Ziller et al. (2013), *** data published by Okae et al. (2014)

		2391_1_hg19		6141_1_hg19		9224_1_hg19		9261_1_hg19		9377_1_hg19		9473_1_hg19	
sample_name	cell/tissue	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.
Monocyte methylome 1 *	monocyte	56%	12	49%	18	58%	13	72%	15	39%	7	79%	12
Monocyte methylome 2 *	monocyte	63%	8	54%	7	31%	12	49%	6	65%	5	80%	5
Frontal_cortex_normal_1 **	cortex	57%	42	32%	46	57%	51	41%	42	63%	49	92%	45
Frontal_cortex_normal_2 **	cortex	47%	32	25%	35	61%	32	45%	28	66%	25	83%	32
Frontal_cortex_AD_1 **	cortex	29%	41	29%	43	69%	43	31%	38	67%	37	82%	41
Frontal_cortex_AD_2 **	cortex	56%	57	35%	54	62%	67	46%	46	67%	47	87%	55
IMR90 **	Lung, foetel, fibroblast	75%	18	24%	12	24%	12	56%	24	44%	15	70%	14
Colon_Primary_Normal **	colon	70%	54	45%	33	45%	33	55%	66	37%	39	79%	44
human sperm **	sperm	2%	6	5%	9	0%	5	6%	6	1%	8	2%	6
adult liver replicate 1 **	liver	87%	63	49%	65	66%	46	62%	57	29%	50	80%	73
adult liver replicate 2 **	liver	71%	49	28%	43	79%	56	48%	69	30%	44	68%	50
hippocampus middle replicate 1 **	hippocampus middle	60%	52	35%	40	62%	41	46%	67	60%	37	90%	60
hippocampus middle replicate 2 **	hippocampus middle	52%	39	33%	35	61%	47	42%	45	57%	38	89%	46
fetal heart (119) **	fetal heart	50%	40	19%	34	66%	40	53%	69	31%	32	66%	40
fetal thymus (1238) **	fetal thymus	45%	41	15%	36	50%	29	44%	57	29%	40	73%	35
fetal adrenal (1244) **	fetal adrenal	26%	25	21%	24	76%	22	28%	29	17%	34	65%	32
fetal muscle leg (1243) **	fetal muscle leg	49%	35	11%	34	61%	25	37%	40	26%	35	77%	35
fetal brain (515) **	fetal brain	58%	24	19%	24	66%	20	39%	28	33%	18	89%	27
human oocytes 1 ***	oocytes	1%	6	5%	7	91%	9	94%	9	2%	8	7%	7
human oocytes 2 ***	oocytes	3%	3	11%	3	100%	2	100%	2	2%	2	23%	2

		13085_1_hg19		13250_1_hg19		15400_1_hg19		15512_1_hg19		16448_1_hg19		16458_1_hg19	
sample_name	cell/tissue	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.
Monocyte methylome 1 *	monocyte	76%	8	22%	10	69%	16	54%	18	79%	13	31%	11
Monocyte methylome 2 *	monocyte	78%	3	1%	4	77%	11	35%	7	90%	10	12%	5
Frontal_cortex_normal_1 **	cortex	83%	47	2%	60	88%	42	86%	28	49%	45	18%	40
Frontal_cortex_normal_2 **	cortex	90%	36	4%	43	91%	29	89%	16	52%	44	11%	33
Frontal_cortex_AD_1 **	cortex	91%	35	6%	47	85%	37	70%	20	68%	38	40%	31
Frontal_cortex_AD_2 **	cortex	91%	42	12%	68	80%	45	83%	27	56%	53	40%	38
IMR90 **	Lung, foetel, fibroblast	91%	21	8%	16	70%	22	80%	21	54%	16	15%	14
Colon_Primary_Normal **	colon	84%	38	7%	44	86%	52	74%	51	88%	61	43%	38
human sperm **	sperm	2%	9	0%	9	3%	6	8%	5	8%	11	2%	8
adult liver replicate 1 **	liver	82%	68	10%	57	69%	69	88%	68	72%	76	31%	43
adult liver replicate 2 **	liver	72%	53	5%	46	70%	49	73%	56	47%	64	37%	47
hippocampus middle replicate 1 **	hippocampus middle	75%	52	10%	48	89%	66	89%	60	56%	59	14%	39
hippocampus middle replicate 2 **	hippocampus middle	72%	35	2%	45	92%	44	83%	36	50%	44	36%	34
fetal heart (119) **	fetal heart	51%	35	5%	37	79%	48	38%	36	31%	44	15%	31
fetal thymus (1238) **	fetal thymus	77%	37	2%	37	68%	35	73%	33	90%	34	25%	24
fetal adrenal (1244) **	fetal adrenal	60%	29	2%	30	69%	27	69%	24	28%	22	20%	21
fetal muscle leg (1243) **	fetal muscle leg	64%	41	1%	38	86%	37	59%	26	37%	31	22%	24
fetal brain (515) **	fetal brain	70%	21	0%	25	90%	27	55%	21	40%	27	20%	24
human oocytes 1 ***	oocytes	0%	2	0%	4	0%	10	94%	6	2%	5	0%	8
human oocytes 2 ***	oocytes	12%	3	1%	3	0%	2	100%	1	2%	2	9%	3

		17031_1_hg19		20403_1_hg19		23548_1_hg19		24982_1_hg19		26859_1_hg19	
sample_name	cell/tissue	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.	meth.	cov.
Monocyte methylome 1 *	monocyte	42%	7	20%	13	59%	14	55%	20	77%	8
Monocyte methylome 2 *	monocyte	46%	18	13%	11	42%	10	55%	12	94%	4
Frontal_cortex_normal_1 **	cortex	65%	40	3%	32	81%	51	27%	44	73%	61
Frontal_cortex_normal_2 **	cortex	57%	31	1%	24	78%	36	34%	35	73%	50
Frontal_cortex_AD_1 **	cortex	35%	30	4%	30	68%	38	29%	38	71%	41
Frontal_cortex_AD_2 **	cortex	52%	51	3%	41	58%	47	35%	59	73%	72
IMR90 **	Lung, foetel, fibroblast	33%	14	12%	9	48%	16	57%	26	61%	18
Colon_Primary_Normal **	colon	72%	46	22%	30	77%	65	55%	63	86%	26
human sperm **	sperm	8%	10	3%	8	3%	9	5%	6	5%	3
adult liver replicate 1 **	liver	64%	56	24%	30	44%	79	58%	94	83%	41
adult liver replicate 2 **	liver	55%	56	16%	32	41%	51	59%	62	79%	33
hippocampus middle replicate 1 **	hippocampus middle	58%	41	5%	38	59%	79	43%	83	88%	41
hippocampus middle replicate 2 **	hippocampus middle	54%	38	7%	30	63%	58	39%	64	88%	27
fetal heart (119) **	fetal heart	41%	38	5%	30	29%	38	51%	53	78%	24
fetal thymus (1238) **	fetal thymus	36%	31	9%	29	40%	31	44%	47	67%	38
fetal adrenal (1244) **	fetal adrenal	27%	25	5%	22	23%	33	46%	36	85%	12
fetal muscle leg (1243) **	fetal muscle leg	16%	36	6%	25	56%	38	45%	44	68%	38
fetal brain (515) **	fetal brain	45%	26	4%	16	31%	30	55%	34	71%	25
human oocytes 1 ***	oocytes	0%	4	7%	6	17%	6	83%	3	4%	9
human oocytes 2 ***	oocytes	13%	1	0%	3	2%	2	89%	5	3%	4

Table S8: Primer sequences for genomic sequence analysis

CGI _ID	**Primer name**	**Primer sequence (5‘-3‘)**	**Tm (°C)**	**Cycles**	**Product size (bp)**	**Betain**	**SNP**
Genomic sequence analysis							
2391_1_hg19	2391_1_hg19_F	GTCTTGCACCAGTCACTCTTG	64	35	272	no	rs11806449
	2391_1_hg19_R	TACTCTACGGCCCTTCACAG					
6141_1_hg19	6141_1_hg19_F	CTGGAATTTTTGGACCATGC	64	45	428	yes	rs1849135
	6141_1_hg19_R	CGCTGGGCTTTCCTCAAG					
9224_1_hg19	9224_1_hg19_F	TTAAAGCAAATCCCAGCCCC	68	35	186	no	rs6959065
	9224_1_hg19_R	AAAAAACAAAGGGCGGTCGC					
9261_1_hg19	9261_1_hg19_F	GACCAAGCTCTCCATTGTCA	63	35	289	no	rs10228640
	9261_1_hg19_R	GGGATCAGGTAGCTGTAGATG					
9377_1_hg19	9377_1_hg19_F	GGCTGGCTCTGGTTAGGTC	64	35	318	no	rs1962367 rs140384299
	9377_1_hg19_R	AGCTCAAAGATGCCAGTGAAG					
9473_1_hg19	9473_1_hg19_F	ACCGCTTTGAAACCCCTACT	62	35	244	no	rs62463194
	9473_1_hg19_R	GGACTTGGGCCTTTGTGTT					
13085_1_hg19	13085_1_hg19_F	CCACCAAAGGGAAGCTTGTA	64	35	319	no	rs10996168
	13085_1_hg19_R	GCTCTGGGCCATGATCTCT					
13250_1_hg19	13250_1_hg19_F	CTCTGAACGCCTGTTCCTTC	64	35	250	no	rs61869800
	13250_1_hg19_R	AGCAGCCTCTCCACTTGGT					
15400_1_hg19	15400_1_hg19_F	ATTGGGCTCAGATCAGGAG	66	35	650	no	rs670519
	15400_1_hg19_R	GAAGAAAAGCCTTTCCCAAG					
15512_1_hg19	15512_1_hg19_F	CCTTCTCACGCAAAGCTTCA	68	35	218	no	rs77023346
	15512_1_hg19_R	GCCCCTAATCTGCCTCCTC					
16448_1_hg19	16448_1_hg19_F	TCCAACACATACTATACCACACG	66	45	234	yes	rs4770130
	16448_1_hg19_R	GATGTTCTCGAGCTCCATGG					
16458_1_hg19	16458_1_hg19_F	GCGTCTGCCCAACTCTTG	64	35	225	no	rs7333053
	16458_1_hg19_R	CTGCCTGCCTGATATGACTG					
17031_1_hg19	17031_1_hg19_F	GCCCCTCCACAAGGAAACT	64	45	512	yes	rs7141760
	17031_1_hg19_R	AAAGGCAGATGGCTCATACC					
20403_1_hg19	20403_1_hg19_F	GGCGGYGCTCTACTTCTG	64	35	256	no	rs11078314
	20403_1_hg19_R	CCATTGGCTTTCACTCAGAC					
23548_1_hg19	23548_1_hg19_F	CCATCTGGCTCAAGCTGTC	68	35	334	no	rs12975447
	23548_1_hg19_R	AAGATGATGGGGGAACTGC					
24982_1_hg19	24982_1_hg19_F	TCATCTTCCGGTTAGCGACA	64	35	226	no	rs45612231
	24982_1_hg19_R	CCAGTTGGCTCTTAATGCCC					
26859_1_hg19	26859_1_hg19_F	CTAGAGCGTCAGGGATCAGG	64	35	279	no	no SNP
	26859_1_hg19_R	GGCGACAAACAAGCGTATCT					

Table S9: Primer sequences for deep bisulfite sequencing (universal tags in bold)

CGI _ID	**Primer name**	**Primer sequence (5‘-3‘)**	**Tm (°C)**	**Product size (bp)**
Deep bisulfite sequencing (universal tags in bold)				
2391_1_hg19	2391_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TGGAGGTTAGTATTTTTTTGATTTG	56	309
	2391_1_hg19_RM13	**CAGGAAACAGCTATGAC**CTAAACAACAAAACCAAACCATATCTA		
6141_1_hg19	6141_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TTGAATTGTTTYGTTAGGGTT	53	360
	6141_1_hg19_RM13	**CAGGAAACAGCTATGAC**ATCTCTTTTCCATTCCATAATTTA		
9224_1_hg19	9224_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TGTTGGGGTATATTTTAAAGTAAATTTTAG	54	258
	9224_1_hg19_RM13	**CAGGAAACAGCTATGAC**AAAACTTAAACAAATYAACAAAAAAAA		
9261_1_hg19	9261_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**ATTTTGATTAAGTTTTTTATTGTTA	50	350
	9261_1_hg19_RM13	**CAGGAAACAGCTATGAC**TCTCCTTCCAAAAATCCTTC		
9473_1_hg19	9473_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TGGATTGAGGAGTGTTGATAGTTAT	56	307
	9473_1_hg19_RM13	**CAGGAAACAGCTATGAC**CCCATAAACTTAAACCTTTATATTTTTATT		
13250_1_hg19	13250_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TTTAGATTAGTTTTTGTTTATTTTGTAG	56	233
	13250_1_hg19_RM13	**CAGGAAACAGCTATGAC**ACTCCAACAACCTCTCCACTTAAT		
15400_1_hg19	15400_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TTGTTTGGTATTTTTAGTTAAATT	50	577
	15400_1_hg19_RM13	**CAGGAAACAGCTATGAC**ATAAAATTATCCAAAAAATTTTCC		
16448_1_hg19	16448_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TGGTTTTTGAGAATATTATGTT	50	342
	16448_1_hg19_RM13	**CAGGAAACAGCTATGAC**ATATARAAAAACTAAAACATCTACC		
16458_1_hg19	16458_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**TGATTAGTTTGTTTGTTTGATATGATTG	48	466
	16458_1_hg19_RM13	**CAGGAAACAGCTATGAC**AAATAATAATAACTCTTTTACCATAACC		
17031_1_hg19	17031_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**GGTTTAGGGGAGAGGTTTAGG	56	294
	17031_1_hg19_RM13	**CAGGAAACAGCTATGAC**CCTTAAAACTATTAATCAAATACTCCAAA		
20403_1_hg19	20403_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**ATTTAAGAAGTGTTTTTTTTGTTT	56	369
	20403_1_hg19_RM13	**CAGGAAACAGCTATGAC**ACTAAATCTTCTACTTCTAAATATCACACT		
23548_1_hg19	23548_1_hg19_Ftag	**CTTGCTTCCTGGCACGAG**GTTAGTAGGGTTTGGTTTTTAGGG	56	312
	23548_1_hg19_RM13	**CAGGAAACAGCTATGAC**CATAAAAATAATAAAAAAACTACCCC		

Table S10: Primer sequences for RNA analysis

Gene	**Primer name**	**Primer sequence (5‘-3‘)**	**Tm (°C)**	**Product size (bp)**
mRNA analysis				
RPS2P32	9261_1_hg19_F	GACCAAGCTCTCCATTGTCA	63	289
	9261_1_hg19_R	GGGATCAGGTAGCTGTAGATG		
IGF2BP3	F_mRNA_IGF2BP3	GCTGAGGAGGAGATCATGAAG	59	378
	R_mRNA_IGF2BP3	GAACTGAGCCTCTGGTGGTC	60	
SNaPshot analysis				
RPS2P32	RPS2P32_SNAPSHOT_F	CACTGCCACCCTGGGCA	66	
	RPS2P32_SNAPSHOT_R	CTTCGCCAACGCCACCTT	65	

Table S11: Detailed description of the analyzed 17 RO-CGIs.

CGI UCSC								METHYLOME PRJEB5800	
CGI_ID	CHR	START	STOP	LENGTH	C+G-CONTENT	RATIO OBS/EXP	NUMBER CGs	MEAN METH	MEAN COVERAGE
2391_1_hg19	chr1	240656253	240656720	467	65%	0,79	39	56%	12
6141_1_hg19	chr4	144833114	144833512	398	54%	1,34	54	49%	18
9224_1_hg19	chr7	16890768	16891087	319	68%	0,7	26	58%	13
9261_1_hg19	chr7	23530434	23530690	256	65%	0,75	20	72%	15
9377_1_hg19	chr7	36010997	36011407	410	76%	0,85	50	39%	7
9473_1_hg19	chr7	52341468	52342266	798	69%	0,62	59	79%	12
13085_1_hg19	chr10	66813635	66814061	426	74%	0,77	44	76%	8
13250_1_hg19	chr10	91596974	91597792	818	67%	0,69	63	22%	10
15400_1_hg19	chr12	31405184	31405545	361	60%	0,79	26	69%	16
15512_1_hg19	chr12	49782965	49783193	228	64%	0,81	19	54%	18
16448_1_hg19	chr13	21893156	21893605	449	67%	0,92	46	79%	13
16458_1_hg19	chr13	23412207	23412623	416	72%	0,83	43	31%	11
17031_1_hg19	chr14	21191657	21191860	203	80%	0,83	27	42%	7
20403_1_hg19	chr17	15686218	15686474	256	72%	0,98	32	20%	13
23548_1_hg19	chr19	21860792	21861016	224	67%	0,89	22	59%	14
24982_1_hg19	chr20	30135076	30135292	216	64%	1,12	25	55%	20
26859_1_hg19	chrX	37026348	37026706	358	64%	0,79	29	77%	8

	RETROCOPY		ALL RETROCOPIES*	REGION -1000BP OF THE RETROCOPY		
CGI_ID	RETROCOPY	EXPRESSION RETROCOPY	TOTAL NUMBER OF RETROCOPIES	-1000bp mean meth	-1000bp mean coverage	-1000bp repeat
2391_1_hg19	*retro_GREM1*	expressed strong	1	68%	15	low complexity
6141_1_hg19	*retro_SAV1*	expressed weak	1	49%	51	no
9224_1_hg19	*retro_ARHGAP20*	expressed weak	2	55%	54	L1
9261_1_hg19	*retro_RPS2*	expressed weak	63	96%	24	ALU
9377_1_hg19	*retro_PPP1R14B*	expressed weak	5	95%	75	simple repeat, L1, hAT-Charlie
9473_1_hg19	*retro_CCDC115*	pseudogene	1	91%	38	low complexity, ALU, ERV1
13085_1_hg19	*retro_NEK4*	pseudogene	7	89%	25	simple repeat, low complexity
13250_1_hg19	*retro_MARK2*	expressed strong noOrf	32	97%	29	ERVL-MaLR, L2
15400_1_hg19	*retro_RPL13AP5*	expressed weak	18	88%	38	simple repeat, ALU, TcMar-Tigger
15512_1_hg19	*retro_FGFR1OP2*	expressed weak	4	86%	47	ALU, L1
16448_1_hg19	*retro_GRK6*	expressed weak	2	92%	42	MIR, L1
16458_1_hg19	*retro_IPMK*	expressed weak noOrf	2	77%	21	ALU, ERVL
17031_1_hg19	*retro_XPO6*	expressed weak	1	96%	45	TcMar-Tigger
20403_1_hg19	*retro_IL6ST*	expressed weak noOrf	1	43%	24	low complexity, ALU, L1, rRNA
23548_1_hg19	*retro_MTDH*	expressed weak	14	54%	65	low complexity
24982_1_hg19	*retro_MCTS1*	expressed strong	1	85%	28	ALU, ERVL-MaLR
26859_1_hg19	*retro_FAM47E*	expressed weak	4	76%	17	L1

	REGION +1000BP OF THE RETROCOPY			ANCESTRAL GENE	
CGI_ID	+1000bp mean meth	+1000bp mean coverage	+1000bp Repeat	ANCESTRAL GENE	ANCESTRAL GENE NUMBER CGIs
2391_1_hg19	87%	28	ALU, MIR	*GREM1*	1
6141_1_hg19	72%	53	ERVL-MaLR, hAT-Charlie	*SAV1*	1
9224_1_hg19	50%	22	simple repeat	*ARHGAP20*	1
9261_1_hg19	84%	27	low complexity, ALU, scRNA, hAT-Charlie	*RPS2*	1
9377_1_hg19	96%	25	MIR, hAT-Charlie	*PPP1R14B*	1
9473_1_hg19	59%	76	L1	*CCDC115*	1
13085_1_hg19	90%	48	ALU	*NEK4*	1
13250_1_hg19	94%	48	L1, hAT-Charlie	*MARK2*	2
15400_1_hg19	94%	59	low complexity	*RPL13AP5*	2
15512_1_hg19	96%	30	ALU, L1	*FGFR1OP2*	1
16448_1_hg19	80%	35	L1	*GRK6*	3
16458_1_hg19	79%	37	ALU, ERVL-MaLR	*IPMK*	1
17031_1_hg19	63%	35	ERV1	*XPO6*	1
20403_1_hg19	55%	31	ALU	*IL6ST*	1
23548_1_hg19	85%	37	simple repeat, L2	*MTDH*	1
24982_1_hg19	88%	20	MIR	*MCTS1*	0
26859_1_hg19	86%	17	ALU	*FAM47E*	1

Table S12: Row data of SNaPshot analysis.

Sample	Peak area	A mat / G pat	ratio	G pat / A mat	ratio
child Trio III DNA G allele	52038	0.40	1	2.47	1
child Trio III DNA A allele	21063				
child Trio III cDNA G allele F	14005	0.34	0.85	2.91	1.18
child Trio III cDNA A allele F	4816				
child Trio III cDNA G allele R	22161	0.23	0.56	4.39	1.78
child Trio III cDNA A allele R	5048				

Interpretation of a Phred Quality Score:

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Strategy of genomic primer design

1. Align all sequences of retrocopies using geneious (http://www.geneious.com/; last accessed 09.01.2016).
2. Download GenBank file of the region of interest including information of SNPs, repeats and exon-intron.-structure (http://www.ensembl.org/; last accessed 09.01.2016) and align to the alignment.
3. Design primers (length around 20bp) manually in regions where are mismatches between the sequences (3' sequence of the primer), and without a SNP inside the primer sequence.
4. Check primers by using the Human BLAT Search (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start; last accessed 09.01.2016).
5. Check primer using UCSC In-Silico PCR (https://genome.ucsc.edu/cgi-bin/hgPcr?hgsid=468504943_qb7OtRiFVIjXoCcd5pnKaAL842Mx; last accessed 09.01.2016) for annealing temperature and single PCR product.
6. Check primer using bisearch (http://bisearch.enzim.hu/?m=genompsearch; last accessed 10.01.2016) for single PCR product.

Strategy of bisulfit primer design

1. Align all bisulfit sequences of retrocopies using geneious (http://www.geneious.com/; last accessed 09.01.2016). For convertion of the genomic sequence to bisulfit we used http://www.zymoresearch.com/tools/bisulfite-primer-seeker (last accessed 10.01.2016)
2. Download GenBank file of the region of interest including information of SNPs, repeats and exon-intron.-structure (http://www.ensembl.org/; last accessed 09.01.2016) and align to the alignment.
3. Design primers (length around 25bp) manually or try to use MethPrimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi; last accessed 10.01.2016) in regions where are mismatches between the sequences (3' sequence of the primer), and without a SNP and CpG.dinucleotides inside the primer sequence.
4. Check primers by using the Human BLAT Search (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start; last accessed 09.01.2016).
5. Check primer using UCSC In-Silico PCR (https://genome.ucsc.edu/cgi-bin/hgPcr?hgsid=468504943_qb7OtRiFVIjXoCcd5pnKaAL842Mx; last accessed 09.01.2016) for annealing temperature.
6. Check primer using bisearch (http://bisearch.enzim.hu/?m=genompsearch; last accessed 10.01.2016) for single PCR product.