

# A Genome-Wide Scan of DNA Methylation Markers for Distinguishing Monozygotic Twins

Qingqing Du,<sup>1</sup> Guijun Zhu,<sup>2</sup> Guangping Fu,<sup>1</sup> Xiaojing Zhang,<sup>1</sup> Lihong Fu,<sup>1</sup> Shujin Li,<sup>1</sup> and Bin Cong<sup>1</sup>

<sup>1</sup>Hebei Key Laboratory of Forensic Medicine, Department of Forensic Medicine, Hebei Medical University, Shijiazhuang, Hebei, China

<sup>2</sup>Intensive Care Unit, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

Identification of individuals within pairs of monozygotic (MZ) twins remains unresolved using common forensic DNA typing technology. For some criminal cases involving MZ twins as suspects, the twins had to be released due to inability to identify which of the pair was the perpetrator. In this study, we performed a genome-wide scan on whole blood-derived DNA from four pairs of healthy phenotypically concordant MZ twins using the methylated DNA immunoprecipitation sequencing technology to identify candidate DNA methylation markers with capacity to distinguish MZ twins within a pair. We identified 38 differential methylation regions showing within-pair methylation differences in all four MZ pairs. These are all located in CpG islands, 17 of which are promoter-associated, 17 are intergenic islands, and four are intragenic islands. Genes associated with these markers are related with cell proliferation, differentiation, and growth and development, including zinc finger proteins, *PRRX2*, *RBBP9*, or are involved in G-protein signaling, such as the regulator of G-protein signaling 16. Further validation studies on additional MZ twins are now required to evaluate the broader utility of these 38 markers for forensic use.

■ **Keywords:** monozygotic twins, DNA methylation, methylated DNA immunoprecipitation, epigenome, CpG island, promoter

At present, genetic markers used for forensic identification mainly include short tandem repeats (STR) and single nucleotide polymorphism (SNP), which can identify any specific individual other than monozygotic (MZ) twins within a pair (Borsting et al., 2014; Hiroaki et al., 2015; Hurth et al., 2010; Kim et al., 2010; Romsos & Vallone, 2015). This is because MZ twins develop from a single fertilized egg and share a near-identical genetic sequence, including the set of STR or SNP markers used in forensic cases. There have been a handful of international criminal cases involving MZ twins internationally. In such cases, the potential perpetrator was released due to inability to definitively identify one of the twins of MZ pair. Therefore, discriminating two MZ individuals is an urgent problem facing the forensic geneticist.

In spite of MZ twins having a near-identical genetic sequence, they nevertheless routinely show discordance for phenotypic traits, such as character, life habits, and disease susceptibility. Much of this variation within MZ twin pairs appears to be related to epigenetic differences (Bell & Spector, 2011; Petronis et al., 2003; Sun et al., 2013). Several studies have now confirmed both whole genome DNA

methylation and locus-specific DNA methylation variation within MZ twin pairs (Baynam et al., 2011; Gordon et al., 2012; Loke et al., 2013; Ollikainen et al., 2010; Poulsen et al., 2007; Sun et al., 2013; Talens et al., 2012; Townsend et al., 2005). Therefore, differences in DNA methylation within MZ twin pairs might provide appropriate biomarkers for differentiating MZ twins.

In the present study, we aimed to determine whether DNA methylation can be used to discriminate MZ twins for forensic use. There are approximately 28 million CpG sites in human genome that could potentially be used for forensic science. To find the appropriate candidate markers, we performed a genome-wide scan for four pairs of phenotypically concordant MZ twins using methylated DNA immunoprecipitation (MeDIP) sequencing technology.

RECEIVED 5 August 2015; ACCEPTED 10 September 2015. First published online 26 October 2015.

ADDRESS FOR CORRESPONDENCE: Shujin Li, Hebei Key Laboratory of Forensic Medicine, Department of Forensic Medicine, Hebei Medical University, 361 Zhongshan Road, Shijiazhuang, Hebei 050017, China. E-mail: shujinli@163.com

**TABLE 1**  
Basic Information of Four Pairs of MZ Twins

No.	Age (years)	Sex	Ethnicity
1	0	Female	Han
2	12	Male	Han
3	24	Female	Han
4	36	Male	Han

## Materials and Methods

### Twins

In accordance with the principle of informed consent, peripheral blood samples were collected from four pairs of MZ twins; each subject signed an informed consent form and filled in the corresponding questionnaire. This study passed the ethical review and was approved by the Hebei Medical University Biomedical Ethics Committee. Brief information about the four pairs of MZ twins is listed in Table 1. Genomic DNA was extracted using QIAamp DNA Blood Kit (Qiagen, Germany) and was genotyped with 19 STR markers using Goldeneye DNA Identification System Basic Kit (Peoplespot, China) to identify the zygosity of twins.

### MeDIP Sequencing

**MeDIP and sequencing library preparation.** MeDIP-sequencing library preparation was performed according to Down et al. (2008) with minor modifications. For MeDIP, genomic DNA was sonicated to ~200–900 bp with a Bioruptor sonicator (Diagenode, Belgium); 800 ng of sonicated DNA was end-repaired, A-tailed, and ligated to single-end adaptors following the standard Illumina genomic DNA protocol. After agarose size selection to remove unligated adaptors, the adaptor-ligated DNA was used for immunoprecipitation using a mouse monoclonal anti-5-methylcytosine antibody (Diagenode, Belgium). For this, DNA was heat-denatured at 94°C for 10 min, rapidly cooled on ice, and immunoprecipitated with 1- $\mu$ L primary antibody overnight at 4°C with rocking agitation in 400- $\mu$ L immunoprecipitation buffer (0.5% BSA in phosphate buffer solution [PBS]). To recover the immunoprecipitated DNA fragments, 200  $\mu$ L of magnetic beads were added and incubated for additional 2 h at 4°C with agitation. After immunoprecipitation, a total of five immunoprecipitation washes were performed with ice-cold immunoprecipitation buffer. A non-specific mouse IgG immunoprecipitation was performed in parallel to methyl DNA immunoprecipitation as a negative control. Washed beads were re-suspended in TE buffer with 0.25% SDS and 0.25 mg/mL proteinase K for 2 h at 65°C and then allowed to cool down to room temperature. MeDIP and supernatant DNA were purified using Qiagen MinElute columns and eluted in 16- $\mu$ L EB (Qiagen, Germany). Fourteen cycles of polymerase chain reaction (PCR) were performed on 5  $\mu$ L of immunoprecipitated DNA using the single-end Illumina PCR primers. The resulting reactions were purified with Qiagen MinElute

columns, after which a final size selection (300–1,000 bp) was performed by electrophoresis in 2% agarose. Libraries were quality-controlled by Agilent 2100 Bioanalyzer. An aliquot of each library was diluted in EB to 5 ng/ $\mu$ L, and 1  $\mu$ L was used in real-time PCR reactions to confirm enrichment for methylated region.

### Sample Quality Control

The enrichment of DNA immunoprecipitation was analyzed by quantitative PCR (qPCR) using a specific methylated site at H19 locus and a non-methylated site at GAPDH locus. The primer sequences for H19 and GAPDH were as follows: H19: F: 5'-GAGCCGCACCAGATCTTCAG-3'; R: 5'-TTGGTGGAAACACACTGTGATCA-3'; GAPDH: F: 5'-CCACAGTCCAGTCCTGGGAACC-3'; R: 5'-GAGCTACGTGCGCCCGTAAAA-3'. Agilent 2100 Bioanalyzer was used for accurate assessment of quality and concentration of sequencing library, from which the size and concentration of each sample was determined after sequencing library preparation.

### DNA Sequencing

The library was denatured with 0.1-M NaOH to generate single-stranded DNA molecules, and loaded onto two channels of flow cell at 8-pM concentration, amplified in situ using TruSeq Rapid SR Cluster Kit (#GD-402-4001, Illumina, USA). Sequencing was carried out by running 100 cycles on Illumina HiSeq 2000 according to manufacturer's instructions.

### Data Analysis

Sequence reads were generated from Illumina HiSeq 2000; image analysis and base calling were performed using Off-Line Basecaller software (OLB V1.8). We obtained 3.6 Gb data, which contained 73 million reads for each sample. After passing Solexa CHASTITY quality filter, the clean reads were aligned to human genome (UCSC HG19) using BOWTIE software (V2.1.0). The number of passed filtering reads of eight samples was from 13.9 million to 64.9 million, and the number of aligned reads was from 12.0 million to 56.7 million, as listed in Table S1 in Supplementary Material.

### Data Normalization and Digital DNA Methylation Profiles (MeDIP-Score)

To quantify the DNA methylation level of any specific region, we used each mapped reads. Methylation score for any region in the genome was defined as the number of reads per kb (Maunakea et al., 2010a, 2010b).

To compare the DNA methylation profiles between multiple samples, the total number of mapped reads of each sample must be normalized by the total number of reference reads (the maximum number of mapped reads of all samples); the raw read counts were normalized accordingly, with an example shown in Table S2 in Supplementary Material.

After normalization, we calculated the MeDIP-score of a specific region according to the length of that region. The DNA methylation status of a specific region was defined as unmethylated if its MeDIP-score was less than 8.31 reads/kb, partially methylated if its MeDIP-score was within 8.31–374.44 reads/kb, and completely methylated if its MeDIP-score is greater than 374.44 reads/kb, with an example is shown in Table S3 in Supplementary Material.

## Results

### CpG island (CGI) Methylation profile

**CpG island (CGI) methylation.** According to the MeDIP score, among 27,841 CGIs, the range of partially methylated, completely methylated, and unmethylated regions of eight samples were 11,324–18,050, 4,189–5,474, and 5,164–11,336 respectively, accounting for 40.67–64.83%, 15.05–19.66%, and 18.55–40.72% of the total CGI regions respectively (Table S4 in Supplementary Material). For all eight samples, about 50% of CpGs located in CGI regions were partially methylated, and the completely methylated regions with MeDIP score higher than 374.44 reads/kb accounted for less than 20% (Table S4). These results indicated that the methylation level in CGIs across the whole genome was not high.

CpG islands are grouped into the following three classes on the basis of their distance to RefSeq genes (Maunakea et al., 2010b): (1) Promoter islands: If an island starts within 1,000-bp upstream of a RefGene transcription start site (TSS), and ends 300-bp downstream of a RefGene transcription start site; (2) Intragenic islands: if an island starts 300-bp downstream of a RefGene transcription start site and ends 300-bp upstream of a RefGene transcription end site; and (3) Intergenic islands: if an island starts 300-bp upstream of a RefGene transcription end site and ends 1,000-bp upstream of neighboring RefGene transcription start site (Figure 1A).

Next, we analyzed the CpG methylation status of different classes of CGIs. The results demonstrated that the ratio of completely methylated promoter CGIs was the lowest, at 1–2%, while the unmethylated regions account for the highest proportion, at 30–60%. In contrast, 35–45% of intragenic CGIs showed complete methylation. The methylation level of intergenic CGIs ranked between promoter CGIs and intragenic CGIs (Figure 1B, and Table S5 in Supplementary Material).

### Promoter Methylation Profile

Of 29,402 promoter regions, partially methylated, completely methylated, and unmethylated regions were 17,213–22,432, 69–265, and 6,880–11,924, accounting for 58.54–76.29%, 0.23–0.90%, and 23.40–40.56% of the total number of regions respectively. This showed that the proportion of completely methylated regions with MeDIP score greater than 374.44 reads/kb was very low, close to zero, and most

promoter regions were in the partially methylated and unmethylated state with lower degree of methylation (Table S4).

We subdivided promoters into the following three classes based on their CpG contents (Mikkelsen et al., 2007): (1) High-CpG-density promoter (HCP): Promoters containing a 500-bp interval from –700 bp to +200 bp with a (G + C)-fraction  $\geq 0.55$  and a CpG observed to expected ratio (o/e)  $\geq 0.6$  classified as HCPs; (2) Low-CpG-density promoter (LCP): Promoters containing no 500-bp interval with CpG o/e  $\geq 0.4$  classified as LCPs; (3) Intermediate-CpG-density promoter (ICP): The remainders that do not fall into either class were classified as ICPs. The results show that the methylation profile of three classes of promoter is similar. Most regions were intermediately methylated, followed by regions with a very low degree of methylation. Complete methylation was rare in all three types of promoters (Figure 1B, and Table S6 in Supplementary Material).

### Gene Body Methylation Profiles

The RefSeq genes that were greater than 3 kb in length were chosen for methylation analysis. The gene body region is defined as +2,000 bp downstream of the transcription start site to the transcription termination site (TTS). In the 24,157 gene body regions, 17,166–20,850 partially methylated, 33–85 completely methylated, and 3,273–6,917 unmethylated regions accounted for 71.06–86.31%, 0.14–0.35%, and 13.55–28.63% of regions respectively (Tables S4 and S7 in Supplementary Material). This indicated that the proportion of completely methylated regions across gene bodies was very low, close to zero, and most gene body regions were in partially methylated and unmethylated states with a lower degree of methylation.

### Identification of Differentially Methylated Regions (DMRs) Within MZ Twin Pairs

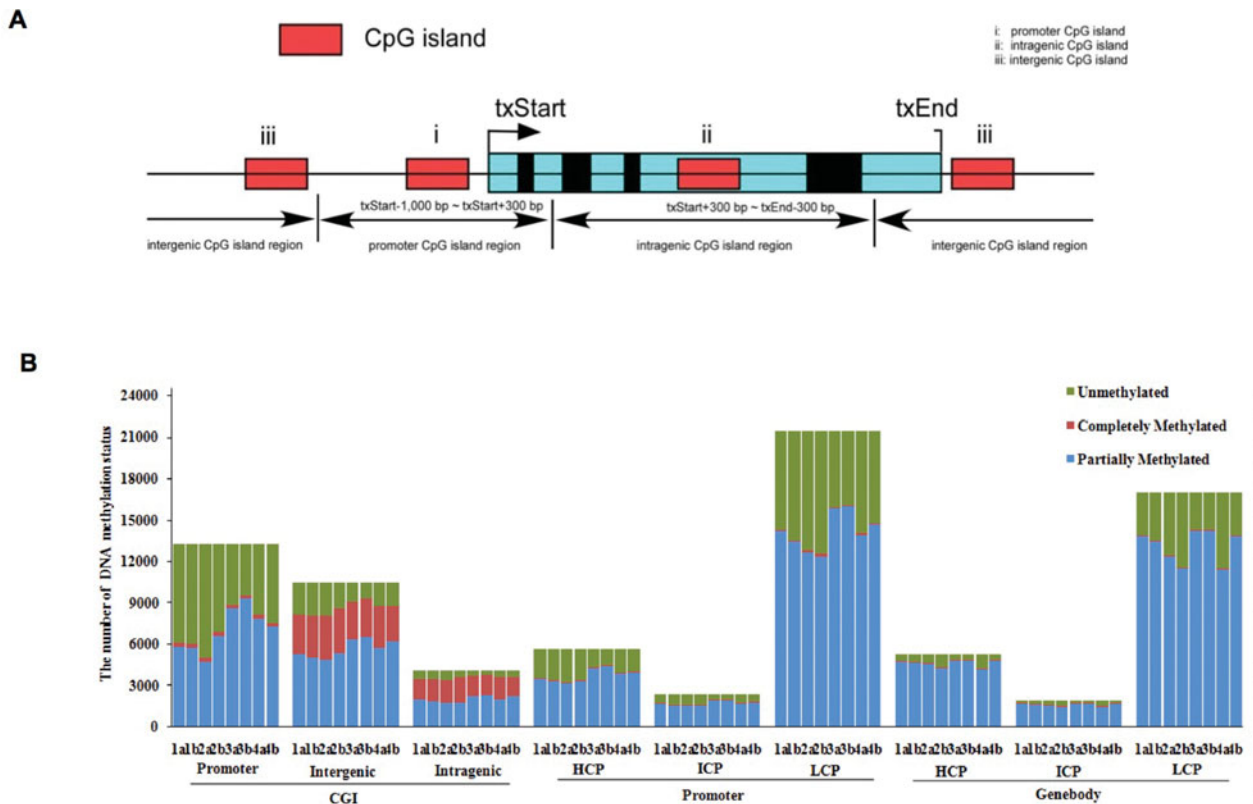
To filter DMRs within twin pairs, fold change (FC) was calculated between the two samples' MeDIP-scores (Table S8 in Supplementary Material). The threshold was  $FC \geq 2.0$  and  $p$  values cut off at  $10^{-4}$  (Shen et al., 2013; Yuan et al., 2014). According to these criteria, a total of 22,889, 21,239, 17,926, and 25,140 DMRs were identified within MZ twin pairs #1, 2, 3, and 4 respectively. Most of the DMRs were located in CGIs, and 12,498, 11,731, 10,229, and 12,998 DMRs within MZ twin pairs #1, 2, 3, and 4 were in CGIs occupying 44.98, 42.13, 36.74, and 46.69% of the total CGI numbers respectively. For promoter regions, MZ twin pairs #1, 2, 3, and 4 had 9,123, 8,362, 6,807, and 10,092 DMRs, accounting for 31.03, 28.44, 23.15, and 34.32% of the total gene promoters respectively. For gene bodies, only 1,268, 1,146, 890, and 2,050 DMRs were observed within four MZ twin pairs, accounting for 5.25, 4.74, 3.68, and 8.49% of such regions respectively (Figure 2, Table 2).

Further analysis of the distribution of DMRs in different types of CGIs showed that DMRs in promoter CGIs

**TABLE 2**

Features of Differentially Methylated Regions (DMRs) of Four MZ Twin Pairs

MZ twin	CGI DMRs	Promoter DMRs	Gene body DMRs	Total DMRs
#1	12,498	9,132	1,268	22,889
#2	11,731	8,362	1,146	21,239
#3	10,229	6,807	890	17,926
#4	12,998	10,092	2,050	25,140

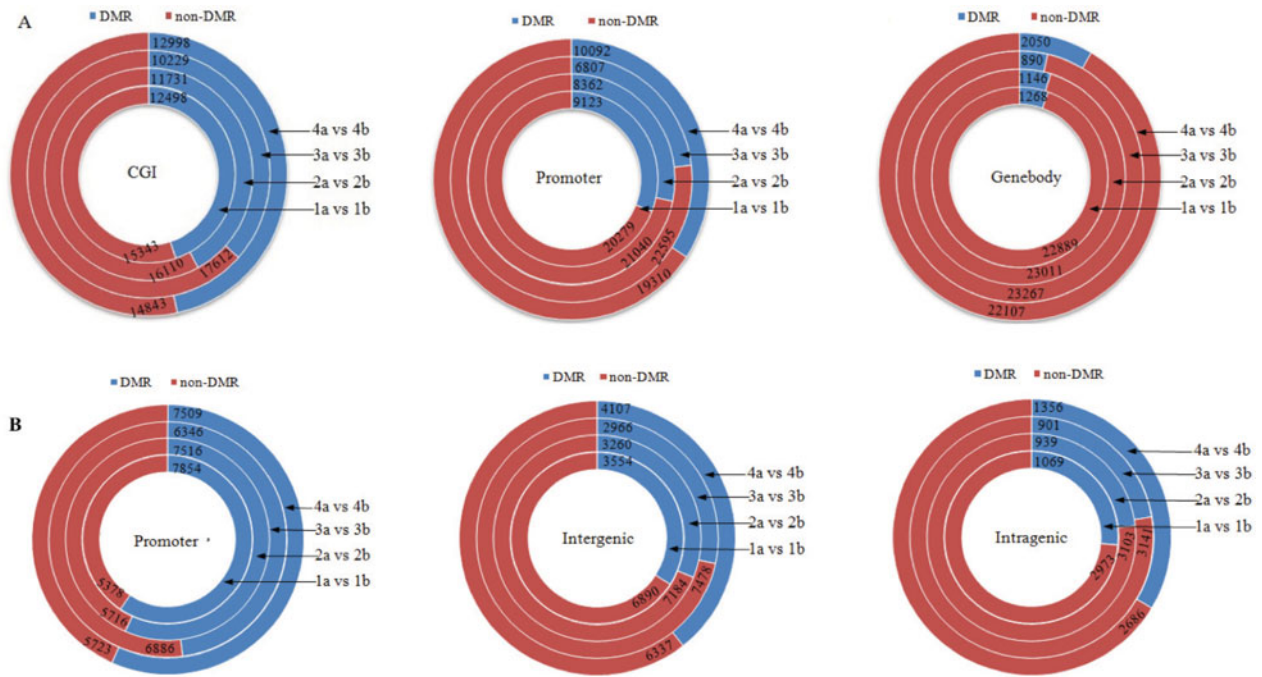
**FIGURE 1**

Whole genome DNA methylation status of eight samples. (A) Three types of CpG islands. (B) Distribution of various DNA methylation status of eight samples across different regions of CpG islands, promoter, and gene body, which were detected by MeDIP. The results showed that the methylation level in promoter CGIs is relatively low, and it is high in the intragenic CGIs. Among the three kinds of promoter regions or gene body, HCP, ICP, and LCP, most of the regions are medium-methylated, followed by regions with very low degree of methylation, and complete methylation is very rare.

were 7,854, 7,526, 6,364, and 7,509 for MZ twin pairs #1, 2, 3, and 4 respectively, accounting for 59.36, 56.80, 47.96, and 56.75% of the total number of promoter CGIs respectively. For intergenic CGIs, each of the four pairs of MZ twins possessed 3,554, 3,260, 2,966, and 4,107 DMRs, occupying 34.03, 31.21, 28.04, and 39.32% of total number of intergenic CGIs respectively. For intragenic CGIs, the number of DMRs within each pair of MZ twin was only 1,069, 939, 901, and 1,356, accounting for 26.45, 23.23, 22.29, and 33.55% of the total intragenic CGIs respectively (Figures 2 and 3). These results suggest that DMRs in CGIs within MZ twin pairs mostly occurred in promoter CGIs, followed by intergenic CGIs, and least often in intragenic CGIs.

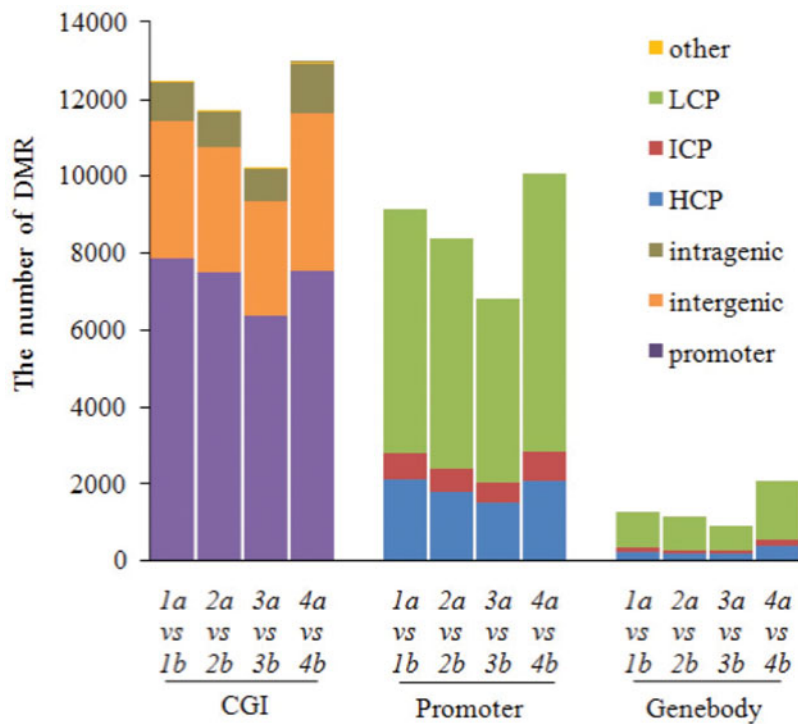
To identify DMRs showing the highest within-pair variation, we conducted a second layer of screening according to two standards: (1) the fold change of MeDIP score was larger than 830, and (2) the MeDIP score of one sample is zero and the other was larger than 8.31. To avoid dividing by zero, 0.01 was added to every count when calculating fold change. Therefore, under this criterion of selection, the selected DMRs must be unmethylated in one sample and partially methylated or completely methylated in another. We called these DMRs filtered through a second layer of screening MZ DMRs (main differentially methylated regions [MDMRs]), meaning DMRs that distinguish within pairs of MZ twins.





**FIGURE 2**

(A) Distribution of DMRs within each MZ twin pair in CGIs, promoters, and gene bodies, and (B) among different regions of CGIs. There were total 22,889, 21,239, 17,926, and 25,140 DMRs in MZ twin pair #1, 2, 3, and 4, respectively. Most of the DMRs were located in CGIs, followed by the promoter regions, and a small number in gene bodies. Further analysis of the distribution of DMRs in different regions of CGIs showed that the DMRs in CGIs within MZ twin pairs mostly occurred in promoter CGIs, followed by intragenic CGIs and CGIs.

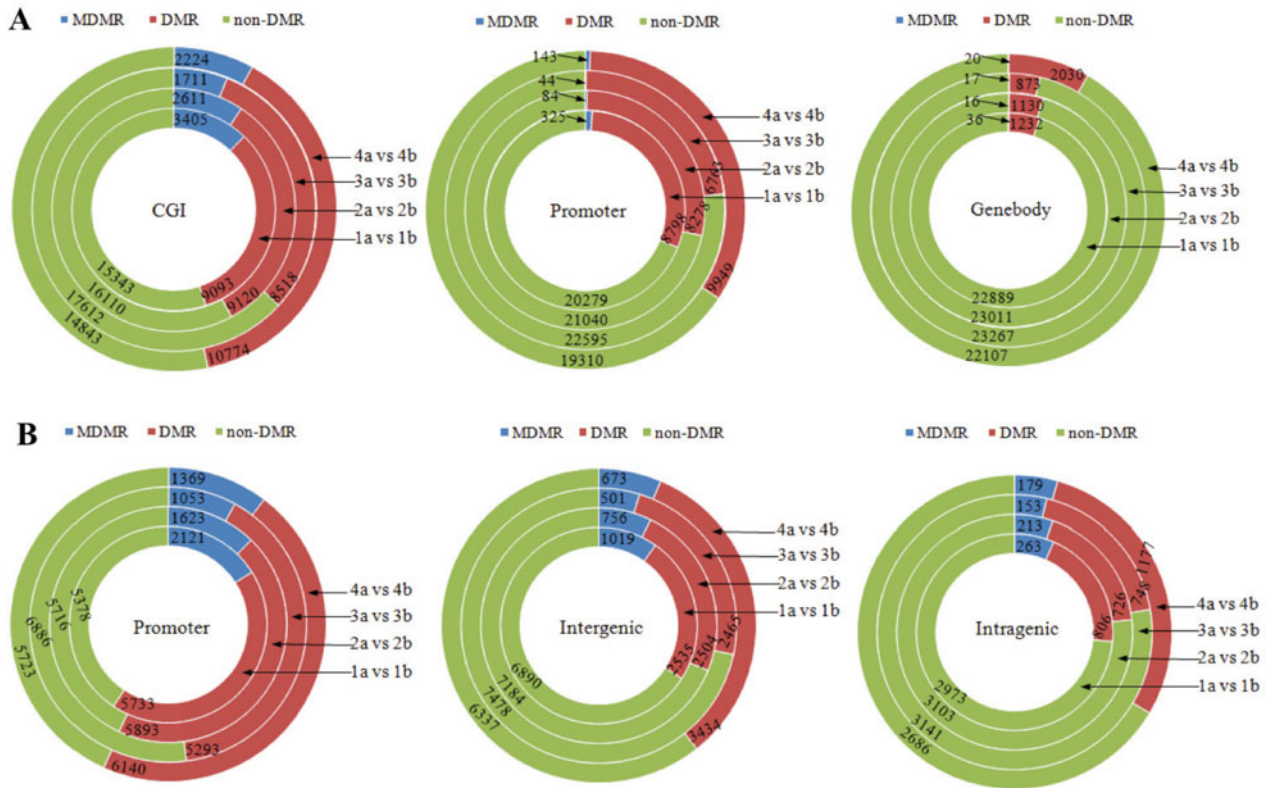


**FIGURE 3**

Comparison of DMRs within each MZ twin pair among different regions of CGIs, promoter and gene bodies.

**TABLE 3**  
Features of Main Differentially Methylated Regions (MDMRs) of Four MZ Twin Pairs

MZ twin	CGI MDMRs	Promoter MDMRs	Gene body MDMRs	Total MDMRs
#1	3,405	325	36	3,766
#2	2,611	84	16	2,711
#3	1,711	44	17	1,772
#4	2,224	143	20	2,387



**FIGURE 4**

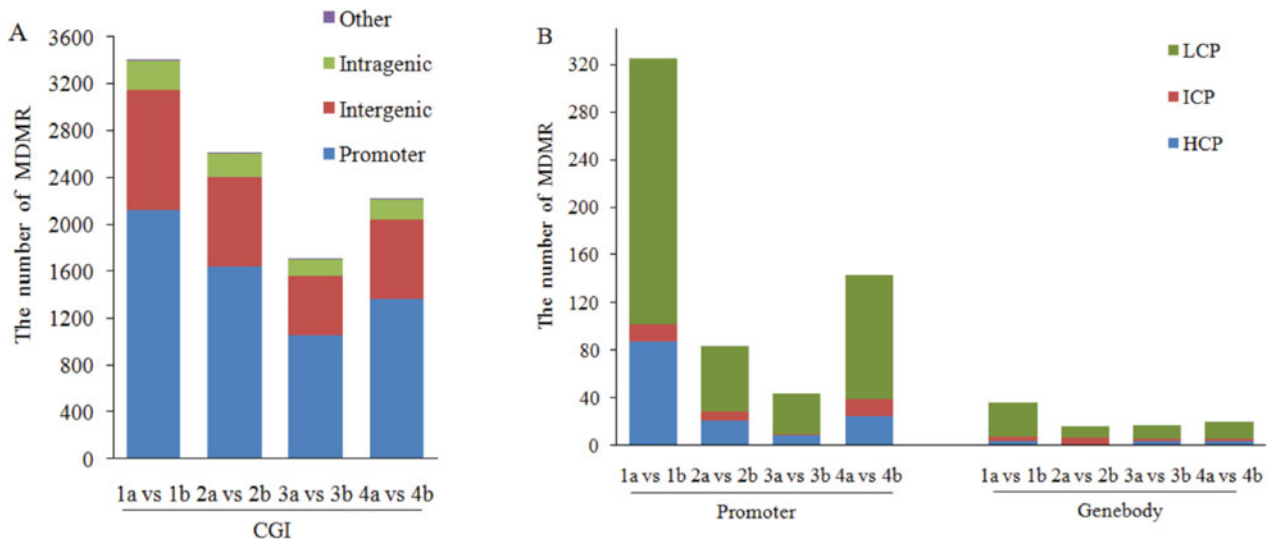
Distribution of MDMRs (A) within each MZ twin pair in CGIs, promoters, and gene bodies, and (B) among different CGI regions. Through the second layer of selection, 3,766, 2,711, 1,772, and 2,387 MDMRs were filtered within MZ twin pair #1, 2, 3, and 4. Most of the MDMRs were located in CGIs, followed by promoter regions, and gene bodies. Further analysis of the distribution of MDMRs in the different regions of CGIs showed that the MDMRs in CGIs within MZ twin pairs mostly occurred in promoter CGIs, followed by intergenic and intragenic CGIs.

Through the second layer of selection, 3,766, 2,711, 1,772, and 2,387 MDMRs were filtered out within MZ twin pairs #1, 2, 3, and 4 respectively. Most MDMRs were located in CGIs, and 3,405, 2,611, 1,711, and 2,224 MDMRs within MZ twin pairs #1, 2, 3, and 4 in CGIs occupy 12.23, 9.38, 6.15, and 7.99% of the total CGIs numbers respectively. For promoter regions, MZ twin pairs #1, 2, 3, and 4 had 325, 84, 44, and 143 MDMRs, accounting for 1.11, 0.29, 0.15, and 0.49% of the total numbers of promoters respectively. For gene bodies, only 36, 16, 17, and 20 MDMRs were observed within four MZ twin pairs, accounting for 0.15, 0.07, 0.07, and 0.08% of the total gene body regions respectively (Figures 4 and 5, Table 3). Further analysis of the distribu-

tion of MDMRs in different types of CGIs showed that the MDMRs in CGIs within MZ twin pair mostly occurred in promoter CGIs, followed by intergenic CGIs, and the least in intragenic CGIs (Figures 4 and 5).

#### MDMRs Consistent Across All Four Pairs of MZ Twins

Thirty-eight MDMRs were shared across all the four pairs of MZ twins, all of which were allocated to CGIs, including 17 promoter-associated, 17 intergenic CGI regions, and 4 intragenic CGI regions. The gene-related MDMRs mainly included genes involved in cell proliferation, differentiation, growth, and development, such as zinc finger proteins

**FIGURE 5**

Comparison of MDMRs within each MZ twin pair among different regions of CGIs, promoters, and gene bodies.

*PRRX2*, *RBBP9*, and genes involved in G-protein signaling, such as the regulator of G-protein signaling *16* (Table 4).

### Comparison of DMRs in Different Ages

To test whether DNA methylation differences become larger with age, we compared DMRs within four MZ twin pairs, who were 0, 12, 24, and 36 years of age. As expected, the number of DMRs was greatest in the 36-year-old MZ twin pair. However, among the remaining three pairs, the newborn MZ twins had the most number of DMRs, followed by the 12-year-old pair, with the 24-year-old pair having the least number of DMRs.

### Discussion

Mammalian genomes are punctuated by DNA sequences containing a high frequency of CpG sites, termed as CGIs. These sequences are characterized with at least 200 bp length, a G + C content of >50%, and an o/e CpG frequency of >0.6 (Illingworth & Bird, 2009). Our results showed that methylation levels in CGIs of all the eight samples were relatively low, especially at promoter CGIs, where about 30–60% of CGIs were unmethylated. In normal somatic cells, the promoter CGIs are typically unmethylated and corresponding genes are frequently expressed. By contrast, these CGIs are aberrantly hypermethylated in cancer, resulting in the suppression of gene expression (Robertson, 2005). However, in intragenic CGIs, complete methylation reached to 35–45%, indicating a higher degree of methylation. The methylation level of intergenic CGIs ranked between promoter and intragenic CGIs. The functional significance of intergenic and intragenic CGIs remains unclear.

Our main focus was on DMRs within MZ twin pairs. Through two layers of screening, we filtered 3,766, 2,711, 1,772, and 2,387 MDMRs within MZ twin pairs 1–4, and most were located within CGIs. Further analysis of the distribution of MDMRs in different types of CGIs showed that MDMRs mostly occurred in promoter CGIs, followed by intergenic CGIs, and then intragenic CGIs. These results indicated that intergenic regions are more likely to have DNA methylation variations than intragenic regions. This is consistent with another study of ours (Fu et al., in press). In that study, we analyzed DNA methylation of two intergenic regions and two genic regions of 119 MZ twin pairs, and found that the intergenic regions showed a higher discriminating efficiency than genic regions, suggesting that intergenic regions might be a better choice to be a candidate region. This may be related with lower selective pressure in intergenic regions than in genic regions.

Among all of the MDMRs, 38 were shared by all four MZ twin pairs and were located in CGIs, including 17 in promoter CGIs, 17 in intergenic CGIs, and 4 in intragenic CGIs. The gene-related MDMRs mainly included genes involved in cell proliferation, differentiation, growth, and development, such as zinc finger protein, *PRRX2*, *RBBP9*, and genes involved in G-protein signaling, such as the regulator of G-protein signaling *16*. These markers need to be further validated using other techniques and repeated on larger samples of MZ twins. Most previous epigenetic studies on MZ twins have selected disease-discordant twins to search for disease-specific epigenetic markers. However, the four pairs of MZ twin that we studied were concordant for general health. These concordant MZ twins shared several DMRs, indicating these DMRs may be representative among MZ twins and valuable markers for discriminating

TABLE 4

## Common Main Differentially Methylated Regions (MDMRs) Across All Four Pairs of MZ Twins

CGI name	CGI_length	CpG number	Relative location	Gene information	FC of 0-year MZ	FC of 12-year MZ	FC of 24-year MZ	FC of 36-year MZ
chr4:4873961–4874173	212	17	Intergenic		11106.17	976.25	5743.35	6799.71
chr10:21805197–21805459	262	22	Intergenic		10306.34	1243.49	2656.13	1266.45
chr13:95359641–95359874	233	31	Intergenic		5151.21	1775.71	5225.8	3558.38
chr11:125774292–125774584	292	36	Promoter	DDX25: NM_013264: 167: -18565	4838.6	2125.18	1787.76	3291.7
chr3:147130342–147130577	235	23	Intragenic	ZIC1: NM_003412: 3279: -4045	4681.85	1386.25	2221.14	1411.84
chr7:100434982–100435193	211	24	Intergenic		4266.4	2315.22	1649.45	6831.93
chr9:116860473–116860695	222	33	Intergenic		4242.98	1467.37	863.15	2165.15
chr1:59281889–59282235	346	40	Intergenic		4083.59	941.85	1107.34	1917.46
chr2:231712775–231712982	207	23	Intergenic		3382.64	999.81	925.62	3204.36
chr14:77422842–77423094	252	31	Intergenic		3175.6	4103.25	2761.5	1907.52
chr16:84001805–84002011	206	22	Promoter	NECAB2: NM_019065: -328: -34469	2913.62	3011.97	7599.06	9329.97
chr19:56154791–56155117	326	31	Promoter	ZNF581: NM_016535: -31: -2034	2889.71	1903.64	3201.82	5895.99
chr12:123450868–123451207	339	36	Promoter	ABCB9: NM_019624: 19: -37498	2778.93	1921.55	1540.04	2363.05
chr16:49309123–49309353	230	24	Intergenic		2609.7	899.93	2269.41	3604.78
chr20:13976700–13977068	368	35	Intragenic	MACROD2: NM_080676: 739: -2056955	2560.02	7304.8	2081.4	871.37
chr1:109791945–109792723	778	109	Promoter	CELSR2: NM_001408: -306: -26043	2421.87	1329.75	2236.36	1492.54
chr17:20811748–20812158	410	34	Intergenic		2297.87	2018.11	1401.47	1172.81
chr5:95768874–95769080	206	18	Promoter	PCSK1: NM_000439: 8: -42938	2286.72	3951.65	1689.46	3888.07
chr14:105293183–105293395	212	22	Intergenic		2222.03	3902.01	4515.08	4692.71
chr17:32484007–32484280	273	29	Promoter	ACCN1: NM_001094: -318: -1144036	2198.8	1515.68	1912.11	6073.3
chr12:115130008–115130242	234	22	Intergenic		2137.75	884.56	2454.81	5476.12
chr20:20433105–20433329	224	17	Intergenic		2103.05	4360.82	2564.35	2961.25
chr2:119599059–119599299	240	20	Intergenic		2084.33	2035.58	10870.45	4145.35
chr1:1072369–1072847	478	55	Intergenic		1971.12	1022.55	1456.32	3122.26
chr12:105351573–105352055	482	60	Intergenic		1954.77	1689.45	2780.63	997.77
chr15:31689500–31689707	207	19	Intragenic	LOC283711: NR_033741: 4558: -7329	1933.37	999.81	5548.74	6190.27
chr12:117348645–117349207	562	78	Promoter	FBXW8: NM_153348: 166: -120027	1676.66	1104.67	1363.26	1140.84
chr6:31620740–31621158	418	29	Promoter	BAT3: NM_080703: -472: -14144;   BAT3: NM_001098534: -779: -14144	1675.64	1484.88	1374.66	1150.38
chr6:82956999–82957507	508	49	Promoter	IBTK: NM_015525: 195: -77298	1575.8	962.22	1712.72	1306.31
chr20:5591490–5591875	385	60	Promoter	GPCPD1: NM_019593: -10: -66602	1224.01	1075.04	1807.87	3015.07
chr9:4491167–4491418	251	20	Intragenic	SLC1A1: NM_004170: 866: -96177	1196.22	1648.44	6236.86	2553.15
chr7:6144047–6144469	422	53	Promoter	USP42: NM_032172: -291: -56936	1116.78	1960.75	1237.34	1572.32
chr9:130158774–130159057	283	22	Promoter	SLC2A8: NM_014580: -549: -11246	1061.07	1726.44	1353.63	1698.67
chr1:182573375–182573760	385	34	Promoter	RGS16: NM_002928: -19: -5809	1039.96	2686.11	2486.68	12480.01
chr4:183062278–183062481	203	20	Intergenic		986.22	4010.03	2829.53	7101.13
chr10:58121059–58121275	216	28	Promoter	ZWINT: NM_032997: -133: -3969	926.93	6029.39	4831.87	2225.27
chr9:132427457–132427679	222	20	Promoter	PRRX2: NM_016307: -351: -57381	901.9	1863.65	5484.69	3734.64
chr20:18477659–18477892	233	18	Promoter	RBBP9: NM_006606: 112: -10586	859.37	888.35	2240.2	3558.38

Note: Ranked by FC between 0-year MZ twin pair. Aligned with UCSC HG19.



MZ twins. A further validation study will be performed using a larger number of samples.

Epigenetic drift refers to age-related changes in epigenome that include those acquired environmentally as well as stochastically (Fraga et al., 2005; Fraga & Esteller, 2007; Jones et al., 2015). Early indications of epigenetic drift were noted in cell culture studies after the observation that clones of a single cell line became epigenetically divergent upon multiple passages (Humpherys et al., 2001). The concept was then used to describe the increase in discordance of DNA methylation between MZ twins as they age (Fraga et al., 2005; Fraga & Esteller, 2007; Martin, 2005). Our data showed that the oldest MZ twin pair (aged 36 years) had the maximum number of DMRs, followed by the newborn pair, then the 12-year-old pair, and the 24-year-old pair had the least number of DMRs. Because of the small number of sample, the correlation analysis between age and DNA methylation differences could not be performed and we were unable to draw firm conclusions about epigenetic drift. However, it is undoubted that there are many DNA methylation differences within newborn MZ twins, indicating that these differences occur during the embryonic development period. Gordon et al. (2012) and Ollikainen et al. (2010) reported that the variable contribution of both intrauterine environmental exposures and underlying genetic factors in the establishment of neonatal epigenome of different tissues confirms the intrauterine period as a sensitive time for the establishment of epigenetic variability in humans. This has implications for the effects of maternal environment on the development of a newborn epigenome, and supports an epigenetic mechanism for the previously described phenomenon of ‘fetal programming’ of disease risk.

When we ranked our 38 MDMRs by fold change, we found that the top gene-related MDMRs location includes zinc finger protein-related genes such as *ZIC1*, *ZNF581*, regulator of G-protein signaling 16 (*RGS16*), N-terminal EF-hand calcium-binding protein 2 (*NECAB2*), and ZW10 interacting kinetochore protein (*ZWINT*). Members of the zinc finger protein family of genes are important during early development. The protein encoded by *RGS16* belongs to the ‘regulator of G-protein signaling’ family. It inhibits signal transduction by increasing the GTPase activity of G-protein alpha subunits. It may also play a role in regulating the kinetics of signaling in the phototransduction cascade. The calcium-binding proteins of the EF-hand super family are involved in the regulation of all aspects of cell function (Grabarek, 2006). The *ZWINT* encoded-protein plays a critical role in cell mitosis and will be changed with cell cycle (Obuse et al., 2004). Therefore, it seems that the genes related to cell growth, development, and mitosis are more likely to be different in DNA methylation between MZ twins.

In conclusion, this study used MeDIP sequencing analysis to screen genome-wide DNA methylation of four pairs

of concordant MZ twins, and a large number of DMRs were found within MZ twin pairs. Most of the DMRs were located in CGIs, especially in promoter CGIs, with a lower number within intragenic regions. The 1,772–3,766 MDMRs with the greatest within-pair differential methylation were further filtered. Finally, 38 MDMRs shared by all four MZ twin pairs were identified as candidate DNA methylation markers for forensically distinguishing MZ twins.

## Acknowledgments

We are grateful to Jeffery M. Craig and Alexandre Lussier for assistance with MeDIP data analysis. This work was supported by the National Natural Science Foundation of China (Shujin Li, grant number: 30973364 and 81373246); Hebei Province Outstanding Youth Science Fund (Shujin Li, grant number: H2012206103); and Hebei Province University Talents Training Program (Shujin Li, grant number: BR2-230).

Shujin Li and Bin Cong designed the study. Qingqing Du and Guangping Fu recruited twins and collected samples. Qingqing Du, Guangping Fu, Xiaojing Zhang, and Lihong Fu completed the experiments; and Qingqing Du and Guijun Zhu carried out the statistical analysis. All authors contributed to the interpretation of data. Shujin Li and Guijun Zhu wrote the first and last drafts of the manuscript, and all the authors made critical revisions. Shujin Li and Bin Cong take responsibility for the integrity of data and accuracy of data analysis.

## Supplementary Material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/thg.2015.73>.

## References

- Baynam, G., Claes, P., Craig, J. M., Goldblatt, J., Kung, S., Le Souef, P., & Walters, M. (2011). Intersections of epigenetics, twinning and developmental asymmetries: Insights into monogenic and complex diseases and a role for 3D facial analysis. *Twin Research and Human Genetics*, 14, 305–315.
- Bell, J. T., & Spector, T. D. (2011). A twin approach to unraveling epigenetics. *Trends in Genetics*, 27, 116–125.
- Borsting, C., Fordyce, S. L., Olofsson, J., Mogensen, H. S., & Morling, N. (2014). Evaluation of the ion Torrent HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Science International: Genetics*, 12, 144–154.
- Down, T. A., Rakyan, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., . . . Beck, S. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, 26, 779–785.
- Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., . . . Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of USA*, 102, 10604–10609.

- Fraga, M. F., & Esteller, M. (2007). Epigenetics and aging: The targets and the marks. *Trends in Genetics*, *23*, 413–418.
- Fu, G.-P., Du, Q.-Q., Zhang, X.-J., Fu, L.-H., Li, S.-J., & Cong, B. (in press). DNA methylation of five target sequences located in genic and intergenic regions for discriminating monozygotic twins. P7, L347.
- Gordon, L., Joo, J. E., Powell, J. E., Ollikainen, M., Novakovic, B., Li, X., . . . Saffery, R. (2012). Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Research*, *22*, 1395–1406.
- Grabarek, Z. (2006). Structural basis for diversity of the EF-hand calcium-binding proteins. *Journal of Molecular Biology*, *359*, 509–525.
- Hiroaki, N., Koji, F., Tetsushi, K., Kazumasa, S., Hiroaki, N., & Kazuyuki, S. (2015). Approaches for identifying multiple-SNP haplotype blocks for use in human identification. *Leg Med (Tokyo)*, *17*(5), 415–420. doi:10.1016/j.legalmed.2015.06.003
- Humpherys, D., Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W. M., 3rd, Biniszkiwicz, D., . . . Jaenisch, R. (2001). Epigenetic instability in ES cells and cloned mice. *Science*, *293*, 95–97.
- Hurth, C., Smith, S. D., Nordquist, A. R., Lenigk, R., Duane, B., Nguyen, D., . . . Zenhausern, F. (2010). An automated instrument for human STR identification: Design, characterization, and experimental validation. *Electrophoresis*, *31*, 3510–3517.
- Illingworth, R. S., & Bird, A. P. (2009). CpG islands — ‘A rough guide’. *FEBS Letters*, *583*, 1713–1720.
- Jones, M. J., Goodman, S. J., & Kobor, M. S. (2015). DNA methylation and healthy human aging. *Aging Cell*. doi:10.1111/accel.12349.
- Kim, J. J., Han, B. G., Lee, H. I., Yoo, H. W., & Lee, J. K. (2010). Development of SNP-based human identification system. *International Journal of Legal Medicine*, *124*, 125–131.
- Loke, Y. J., Novakovic, B., Ollikainen, M., Wallace, E. M., Umstad, M. P., Permezel, M., . . . Craig, J. M. (2013). The peri/postnatal epigenetic twins study (PETS). *Twin Research and Human Genetics*, *16*, 13–20.
- Martin, G. M. (2005). Epigenetic drift in aging identical twins. *Proceedings of the National Academy of Sciences of USA*, *102*, 10413–10414.
- Maunakea, A. K., Chepelev, I., & Zhao, K. (2010a). Epigenome mapping in normal and disease states. *Circulation Research*, *107*, 327–339.
- Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D’Souza, C., Fouse, S. D., . . . Costello, J. F. (2010b). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, *466*, 253–257.
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., . . . Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*, 553–560.
- Obuse, C., Iwasaki, O., Kiyomitsu, T., Goshima, G., Toyoda, Y., & Yanagida, M. (2004). A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. *Nature Cell Biology*, *6*, 1135–1141.
- Ollikainen, M., Smith, K. R., Joo, E. J., Ng, H. K., Andronikos, R., Novakovic, B., . . . Craig, J. M. (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Human Molecular Genetics*, *19*, 4176–4188.
- Petronis, A., Gottesman, I. I., Kan, P., Kennedy, J. L., Basile, V. S., Paterson, A. D., . . . Popendikyte, V. (2003). Monozygotic twins exhibit numerous epigenetic differences: Clues to twin discordance? *Schizophrenia Bulletin*, *29*, 169–178.
- Poulsen, P., Esteller, M., Vaag, A., & Fraga, M. F. (2007). The epigenetic basis of twin discordance in age-related diseases. *Pediatric Research*, *61*, 38R–42R.
- Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics*, *6*, 597–610.
- Romsos, E. L., & Vallone, P. M. (2015). Rapid PCR of STR markers: Applications to human identification. *Forensic Sci Int Genet*, *18*, 90–99. doi:10.1016/j.fsigen.2015.04.008
- Shen, L., Shao, N. Y., Liu, X., Maze, I., Feng, J., & Nestler, E. J. (2013). DiffReps: Detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One*, *8*, e65598.
- Sun, C., Burgner, D. P., Ponsonby, A. L., Saffery, R., Huang, R. C., Vuillermin, P. J., . . . Craig, J. M. (2013). Effects of early-life environment and epigenetics on cardiovascular disease risk in children: Highlighting the role of twin studies. *Pediatric Research*, *73*, 523–530.
- Talens, R. P., Christensen, K., Putter, H., Willemsen, G., Christiansen, L., Kremer, D., . . . Heijmans, B. T. (2012). Epigenetic variation during the adult lifespan: Cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell*, *11*, 694–703.
- Townsend, G. C., Richards, L., Hughes, T., Pinkerton, S., & Schwerdt, W. (2005). Epigenetic influences may explain dental differences in monozygotic twin pairs. *Australian Dental Journal*, *50*, 95–100.
- Yuan, W., Xia, Y., Bell, C. G., Yet, I., Ferreira, T., Ward, K. J., . . . Spector, T. D. (2014). An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. *Nature Communications*, *5*, 5719.