Sharing a Placenta is Associated With a Greater Similarity in DNA Methylation in Monochorionic Versus Dichorionic Twin Pars in Blood at Age 14

Masato Bui,[1] Beben Benyamin,[1] Sonia Shah,[1] Anjali K. Henders,[1] Nicholas G. Martin,[2] Grant W. Montgomery,[2] and Allan F. McRae[1]

[1]*The Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia*
[2]*QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia*

Monozygotic (MZ) twins provide a natural system for investigating developmental plasticity and the potential epigenetic origins of disease. A major difference in the intrauterine environment between MZ pairs is whether they share a common placenta or have separate placentas. Using DNA methylation measured at >400,000 points in the genome on the Illumina HumanMethylation450 array, we demonstrate that the co-twins of MZ pairs (average age of 14) that shared a common placenta ($n = 18$ pairs) have more similar DNA methylation levels in blood throughout the genome relative to those with separate placentas ($n = 16$ pairs). Functional annotation of the genomic regions that show significantly different correlation between monochorionic (MC) and dichorionic (DC) MZ pairs found an over-representation of genes involved in the regulation of transcription, neuronal development, and cellular differentiation. These results support the idea that prenatal environmental exposures may have a lasting effect on an individual's epigenetic landscape, and the potential for these changes to have functional consequences.

∎ **Keywords:** DNA methylation, monozygotic twins, chorionicity

Studies of twins have made substantial contributions to the understanding of the relative importance of genes and the environment in many human traits (Polderman et al., 2015). MZ twins are particularly useful in the investigation of developmental plasticity, as the co-twins are essentially genetically identical, and are matched for age, sex, and cohort-specific effects (Bell & Spector, 2011). The observation of epigenetic discordance within MZ twin pairs could support an epigenetic mechanism for the developmental origins of health and disease (DOHaD) paradigm where intrauterine exposures predispose an individual to disease later in life (Gordon et al., 2012; Loke et al., 2013).

MZ twins can be categorized by their type of placentation. Whereas MC pairs share the same placenta, individuals in a DC pair each have their own placenta. The type of placentation is determined by the timing of embryo splitting post conception. It is thought that twinning that occurs early (usually within the first four days), results in DC or separate placentas, similar to dizygotic twins. When twinning occurs at a later stage, the twin pair may be MC.

Given the important role in nutrient transfer and buffering against potentially harmful in utero exposures, it would seem a reasonable assumption that pairs of MZ twins that share a single placenta may be more similar epigenetically compared to those with separate placentas. However, Kaminsky et al. (2009) found that the 10 pairs of DC twin pairs were more similar than 10 MC twin pairs for DNA methylation from buccal cells, with the mean of the per-probe intra-class correlations being close to zero for MC twin pairs. A different approach to measuring similarity of MZ twin pairs was taken in Gordon et al. (2012), who looked at distributions of intra-class correlations take over all probes within a MZ pair as well as Euclidean distances between members of an MZ pair for both cord blood mononuclear cells and human umbilical vascular endothelia cells. As in Kaminsky et al. (2009), this study provides an indication that MC twin pairs were less similar than DC MZ twin pairs, but reducing the data to a single correlation measure

per twin pair (eight MC, nine and five DC) means these differences were not statistically significant.

Using data from the Brisbane Systems Genetics Study (BSGS; McRae et al., 2014), we investigated the role of placentation on the similarity of DNA methylation levels of adolescent twin pairs. We show that MC MZ twin pairs are more epigenetically similar. Regions of the genome showing significant differences in within-pair similarity in DNA methylation between MC and DC MZ pairs are identified and demonstrated to be involved in regulation of transcription.

Methods

Study Participants

The BSGS is described in detail elsewhere (McRae et al., 2014; Powell et al., 2012). In brief, the BSGS has DNA methylation measured on 614 individuals from 117 families of European descent recruited as part of a study on adolescent twins. The mothers of these families were asked on multiple occasions whether their twins had 'separate placentas', 'joined placentas', or a 'single placenta'. Taking only twin pairs that were consistently reported to have either a single or two placentas left 18 MC (10 male pairs and 8 female, average age 13.6 years, *SE* 1.9), and 16 DC twin pairs (8 male pairs, 8 female, average age 18.8 years, *SE* 2.1).

DNA Methylation

Extraction of DNA and bisulphite conversion is described in detail elsewhere (McRae et al., 2014). Bisulfite converted DNA samples were hybridized to Illumina HumanMethylation450 BeadChips using the Infinium HD Methylation protocol and Tecan robotics (Illumina, San Diego, CA, USA). The HM450 BeadChip-assessed methylation status was interrogated at 485,577 CpG sites across the genome. It provides coverage of 99% of RefSeq genes. Methylation scores for each CpG site are obtained as a ratio of the intensities of fluorescent signals and are represented as β values. Samples in the BSGS study were randomly placed with respect to the chip they were measured on and to the position on that chip in order to avoid any confounding with family. In particular, the design assured that no twin pair had both members on the same array. DNA methylation data are available at the Gene Expression Omnibus under accession code GSE56105.

Probes that have been annotated as binding to multiple chromosomes (Price et al., 2013) were removed from the analysis, as were non-CpG probes. In addition, all probes on the sex chromosomes were excluded as these are not comparable across the sexes due to high methylation in females on the X-chromosome from X inactivation and the lack of Y-chromosome in females. The probability of a probe within a sample either being called as missing or with a detection *p* value less than .001 was estimated from the average rate across all probes and samples. A threshold for probes

showing significant deviation from random missingness (or excess poor binding) was determined by testing against a binomial distribution for the number of samples at the 0.05 significance level with a Bonferroni correction for the number of probes. Probes with more than 11 individuals with missing data or more than five individuals with detection *p* values >.001 were removed. After cleaning, data for a total of 417,069 probes remained for subsequent analysis.

Full details of normalization are described elsewhere (McRae et al., 2014). In brief, individual probes were normalized across all samples using a generalized linear model with a logistic link function using the entire BSGS dataset. Corrections were made for the effects of chip (which encompasses batch processing effects) position on the chip, sex, age, age^2, and interactions between sex and age, and blood cell counts estimated from the methylation data (Houseman et al., 2012).

Testing Differences in MZ Correlations

For each of the 417,069 DNA methylation probes, the MZ pair intra-class correlation was estimated for both MC and DC pairs. Then the test-statistic (*z* score) for the difference between two correlations was calculated as

$$z = \frac{\left(\frac{1}{2}\ln\frac{1+r_1}{1-r_1}\right) - \left(\frac{1}{2}\ln\frac{1+r_2}{1-r_2}\right)}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}.$$

Functional Categorization of Genes

The gene with its transcription start site nearest to the array probe was identified for all probes (Price et al., 2013). Due to the indication of an inflation of probes with low *p* values in the QQ-plot from the MWAS, we considered all probes with a *p* value less than 10^{-4} for the gene set analysis. The 333 unique genes from 351 probes were tested for over-representation of functional annotation using DAVID — The Database for Annotation, Visualization, and Integrated Discovery — v6.7 (Huang et al., 2009a, 2009b).

Results

Figure 1 shows the distribution of intra-class correlations from the 18 MC and 16 DC twin pairs. For example, the MC twin pairs have significantly more probes with correlations greater than 0.5 (19.4% vs. 14.9%, $z = 54$, *p* value smaller than machine precision).

While the power in this study comes from the comparison of the distribution of intra-class correlations between MC and DC MZ twin pairs, we also performed a methylome-wide association study (MWAS), where each methylation probe across the genome is tested for having a difference in correlation between MC and DC MZ twin pairs. Despite its low power, the MWAS identified a number of significant genomic probes (Figure 2a). The eight probes passing a stringent Bonferroni significance threshold are
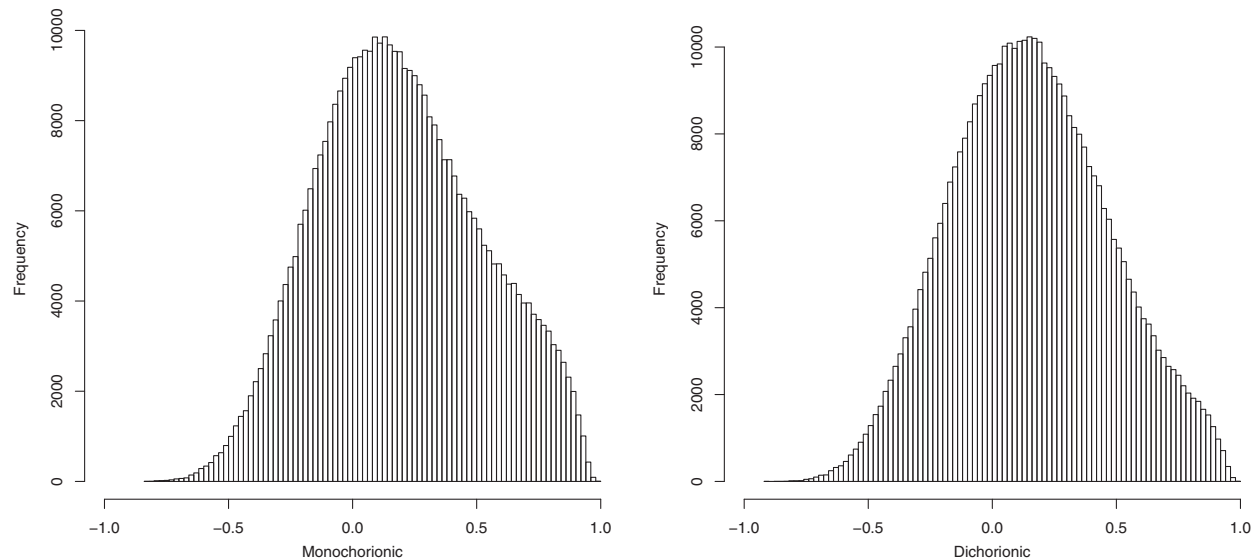
FIGURE 1

Correlations between MZ twin pairs that were MC (left, *n* = 18 pairs) and DC (right, *n* = 16). The MC MZ twins have a larger number probes with correlations greater than 0.5.

TABLE 1

Summary of Probes Showing Significantly Different (Bonferroni Corrected) Correlations Within MZ pairs That Shared or Had Separate Placentas

Probe	Chr	BP*	Annotation	*p* value
cg08737094	6	26,569,535	Intergenic	8.0×10^{-8}
cg10507304	6	27,637,520	Intergenic	1.0×10^{-7}
cg17297045	12	125,581,343	Intronic (*AACS*)	2.5×10^{-8}
cg17327115	8	20,112,950	Upstream (*LZTS1*)	8.6×10^{-8}
cg22557003	4	62,774,948	Intronic (*LPHN3*)	1.4×10^{-8}
cg24022528	7	140,393,175	Intronic (*ADCK2*)	8.3×10^{-9}
cg25548582	13	52,598,798	Exonic (*ALG11*)	1.6×10^{-9}
cg27073561	17	8,045,617	Exonic (*PER1*)	6.3×10^{-9}

Note: *Position of probe CpG target from genome build hg19.

summarized in Table 1. Two of these probes are located near the MHC region on chromosome 6, which also contains a large number of probes with suggestive ($p < 10^{-5}$) significance (cg10507304, $p = 1.0 \times 10^{-7}$; ch18113826, $p = 39 \times 10^{-7}$; ch14047561, $p = 1.2 \times 10^{-6}$; ch11040238, $p = 2.7 \times 10^{-6}$; ch01792601, $p = 9.4 \times 10^{-6}$). The Q-Q plot (Figure 2b) shows a deviation from the expected line for large test statistics (lambda = 1.09).

Given the deviation of the QQ plot from expected under the null, we investigated the functionality of the 351 probes with a *p* value less than 10^{-4} as this will capture additional differences between MC and DC twins at the cost of some false positives. A functional classification of the 333 unique genes with the closest transcription start site to the probes using DAVID found a number of general biological processes significantly over-represented with a Benjamini–Hochberg *p* value < .001 (Table 2). The significantly over-represented categories were primarily involved in the regu-

lation of transcription, but also included categories related to neuronal development and cellular differentiation. The probes with *p* value less than 10^{-4} also show a significantly different ($p = .01$) distribution of HIL annotation (Price et al., 2013), with less probes in high-density CpG islands (24% vs. 32% expected) and more probes in intermediate density areas (30% vs. 24% expected). Genomic annotation (e.g., exonic, intronic, upstream, intergenic) of these CpG sites using ANNOVAR (Wang et al., 2010) found no significant deviation for the 351 probes from the total set ($p = .94$).

Discussion

This study has demonstrated that the blood of co-twins from MZ twin pairs sharing a common placenta are more epigenetically similar than those with separate placentas in blood samples collected when the twins were on average 14 years old. This effect was observed in adolescent twin pairs, indicating that epigenetic changes due to environmental exposures in utero remain throughout early life. This result is in contrast to an earlier study (Kaminsky et al., 2009), which found a higher correlation in DC twin pairs. However, they also found an average MC MZ correlation of close to zero, which is inconsistent with our understanding of the genetic control of DNA methylation (McRae et al., 2014), and indicative of issues due to potential batch effects in their study. Gordon et al. (2012) took a different approach to measuring similarity in MZ twin pairs, which generated a single correlation (or Euclidian distance) across all probes in a twin pair. This has the disadvantage that only a single measure is generated per twin pair and the power afforded by using ~400,000 comparisons across the genome is lost, which
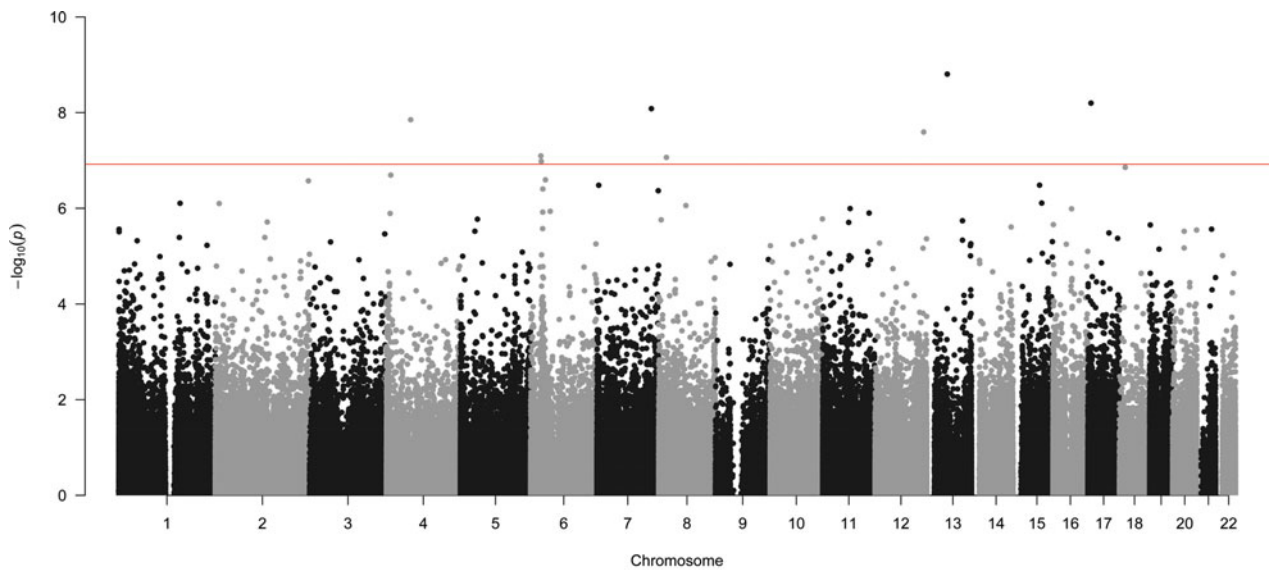
(a)
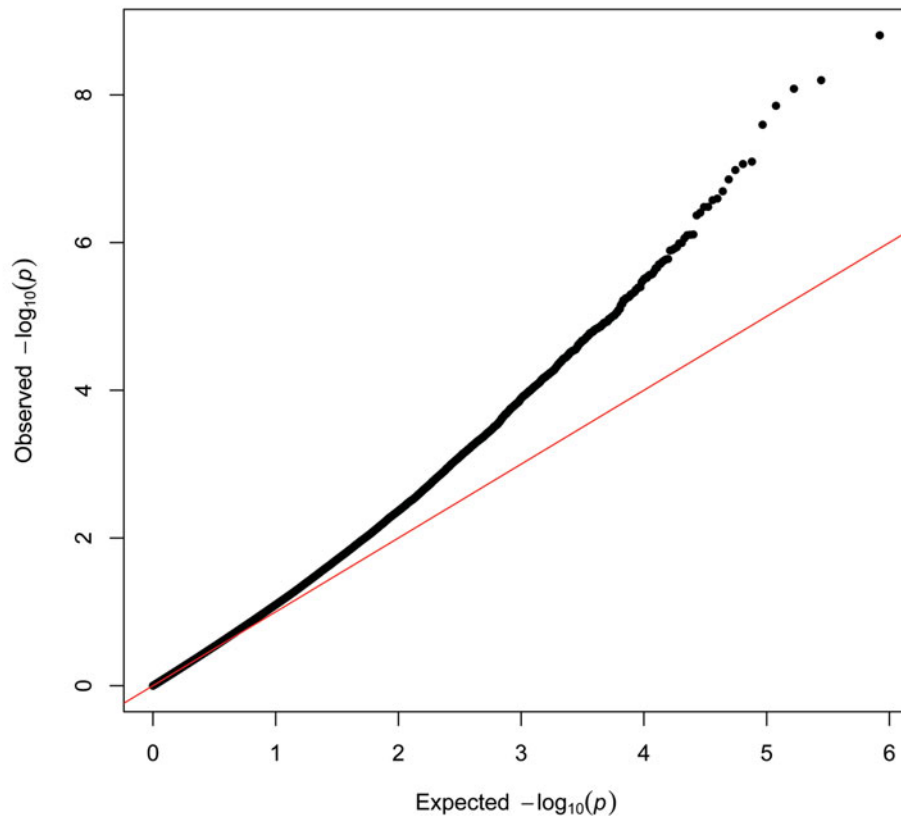
(b)

FIGURE 2

(Colour online) (a) Manhattan plot of the test for the difference MZ twin pair correlations in DNA methylation between MC and DC pairs. A Bonferroni genome-wide significance is indicated. (b) QQ-plot showing an inflation of significant test statistics beyond expected by chance.

TABLE 2

Significantly Over-Represented Gene-Ontology Categories for Methylation Probes with Test Statistics $p < 10^{-3}$

DAVID category	Term	p value (Benjamini–Hochberg)
GOTERM_MF_FAT	Transcription factor activity	6.7×10^{-4}
GOTERM_BP_FAT	Neuron differentiation	4.4×10^{-3}
GOTERM_BP_FAT	Regulation of transcription, DNA dependent	5.0×10^{-3}
GOTERM_BP_FAT	Regulation of RNA metabolic process	6.0×10^{-3}
GOTERM_BP_FAT	Neuron development	7.5×10^{-3}
SP_PIR_KEYWORDS	DNA-binding	8.1×10^{-3}
GOTERM_BP_FAT	Regulation of transcription	8.8×10^{-3}
GOTERM_BP_FAT	Cell morphogenesis involved in differentiation	8.8×10^{-3}
SP_PIR_KEYWORDS	Homeobox	9.0×10^{-3}
GOTERM_MF_FAT	Transcription regulator activity	9.4×10^{-3}

resulted in their study showing no significant differences between MC and DC twins.

An investigation of the potential function of those regions that are more correlated in MC than DC MZ twins identified a number of gene ontology categories that were significantly over-represented. These included low level biological processes involved in the regulation of transcription, neuronal development, and cellular differentiation. A reasonable extrapolation of this result is that MC MZ twins would be more phenotypically similar for complex traits and disease status than DC MZ twins. However, there is conflicting evidence for this in the literature, with MC MZ twins found to be more concordant than DC MZ twins for intelligence (Jacobs et al., 2001), but more discordant for birth weight (Carroll et al., 2005). Indeed, a review by Charlemain and Pons (1998) concludes that MC MZ are generally more similar that DC MZ for 'behavioral' phenotypes and less similar for 'biological' phenotypes. For birth weight it is assumed the increased discordance for MC MZ twins is due to greater competition for the shared blood supply in the placenta for most MC twins. This is a potential explanation for the increased discordance in other biological phenotypes, but it is unclear why it would not also apply to behavioral phenotypes as well.

The difference in similarity between MC and DC MZ twins has implications for the estimation of heritability using twin models. Dizygotic twins have separate placentas, and so have an intra-uterine environment equivalent to only DC MZ twins, so it is possible that the common environment assumption when estimating heritabilities would only hold when excluding MC MZ twins. This study was limited in only comparing 18 and 16 pairs of MC and DC MZ twins, and the deviation from the expected line on the Q-Q plot indicates that we would find many more significant regions with a larger cohort. In addition, using clinically derived chronicity would have the potential to increase the separation between MC and DZ twins, as many DC placentas are fused and essentially look like a single placenta. However, it is of note that the average correlation of DNA methylation levels between a range of relative pairs from the complete BSGS dataset — of which the individuals analyzed

here are a subset — fitted remarkably well to a simple additive genetic model and any common environmental effect for twin pairs being small. Moreover, how this differential methylation translates into differential measured phenotypes is unknown, as is whether the same effect is present for other tissues.

Overall, this study supports previous findings implicating the prenatal environment in mediating specific epigenome signatures that remain through to at least adolescence (Richmond et al., 2015). This supports the idea of the importance of a positive prenatal environment due to its potential role in preventing disease through epigenetic dysregulation (Perera & Herbstman, 2011).

Acknowledgments

References

Bell, J. T., & Spector, T. D. (2011). A twin approach to unraveling epigenetics. *Trends in Genetics*, *27*, 116–25.

Carroll, S. G. M, Tyfield, L., Reeve, L., Porter, H., Soothill, P., & Kyle, P. M. (2005). Is zygosity or chorionicity the main determinant of fetal outcome in twin pregnancies. *American Journal of Obstetrics and Gynecology*, *193*, 757–761.

Charlemain, C., & Pons, J.-C. (1998). What monozygotic twins tell us about genetics determinism. *Race Gender and Class*, *5*, 12–40.

Gordon, L., Joo, J. E., Powell, J. E., Ollikainen, M., Novakovic, B., Li, X., . . . Saffery, R. (2012). Neonatial 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.

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., . . . Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*, 86.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*, 1–13.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*, 44–57.

Jacobs, N., Van Gestel, S., Derom, C., Thiery, E., Vernon, P., Derom, R., & Vlietinck, R. (2001). Heritability estimates of intelligence in twins: Effect of chorion type. *Behavior Genetics*, *31*, 209–217.

Kaminsky, Z. A., Tang, T., Wang, S.-C., Ptak, C., Oh, G. H. T., Wong, A. H. C., . . . Petronis, A. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics*, *41*, 240–245.

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.

McRae, A. F., Powell, J. E., Henders, A. K., Bowdler, L., Hemani, G., Shah, S., . . . Montgomery, G. W. (2014). Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biology*, *15*, R73.

Perera, F., & Herbstman, J. (2011). Prenatal environmental exposures, epigenetics, and disease. *Reproductive Toxicology*, *31*, 363–373.

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*, 702–709.

Powell, J. E., Henders, A. K., McRae, A. F., Caracella, A., Smith, S., Wright, M. J., . . . Montgomery, G. W. (2012). The brisbane systems genetics study: Genetical genomics meets complex trait genetics. *PLoS One*, *7*, e35430.

Price, E. M., Cotton, A. M., Lam, L. L., Farré, P., Emberly, E., Brown, C. J., Robinson, W. P., & Kobor, M. S. (2013). Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics & Chromatin*, *6*, 4.

Richmond, R. C., Simpkin, A. J., Woodward, G., Gaunt, T. R., Lyttleton, O., McArdle, W. L., . . . Relton, C. L. (2015). Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: Findings from the avon longitudinal study of parents and childern (ALSPAC). *Human Molecular Genetics*, *24*, 2201–2217.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Research*, *38*, e164.