

RESEARCH

Open Access

Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance

David Martino¹, Yuk Jin Loke^{2,3}, Lavinia Gordon⁴, Miina Ollikainen⁵, Mark N Cruickshank^{2,6}, Richard Saffery^{1,3†} and Jeffrey M Craig^{2,3*†}

Abstract

Background: The extent to which development- and age-associated epigenetic changes are influenced by genetic, environmental and stochastic factors remains to be discovered. Twins provide an ideal model with which to investigate these influences but previous cross-sectional twin studies provide contradictory evidence of within-pair epigenetic drift over time. Longitudinal twin studies can potentially address this discrepancy.

Results: In a pilot, genome-scale study of DNA from buccal epithelium, a relatively homogeneous tissue, we show that one-third of the CpGs assayed show dynamic methylation between birth and 18 months. Although all classes of annotated genomic regions assessed show an increase in DNA methylation over time, probes located in intragenic regions, enhancers and low-density CpG promoters are significantly over-represented, while CpG islands and high-CpG density promoters are depleted among the most dynamic probes. Comparison of co-twins demonstrated that within-pair drift in DNA methylation in our cohort is specific to a subset of pairs, who show more differences at 18 months. The rest of the pairs show either minimal change in methylation discordance, or more similar, converging methylation profiles at 18 months. As with age-associated regions, sites that change in their level of within-pair discordance between birth and 18 months are enriched in genes involved in development, but the average magnitude of change is smaller than for longitudinal change.

Conclusions: Our findings suggest that DNA methylation in buccal epithelium is influenced by non-shared stochastic and environmental factors that could reflect a degree of epigenetic plasticity within an otherwise constrained developmental program.

Background

Epigenetic modifications such as DNA methylation play an important role in development, ageing and disease [1-3]. However, the factors that influence epigenetic dynamics are poorly understood. Twin studies have the potential to estimate genetic components of epigenetic state [4,5] and have demonstrated that gene expression and DNA methylation profiles can both be influenced by allelic, stochastic and environmental factors [6-10].

Non-shared environmental and stochastic factors together have been estimated to be the largest influence on promoter methylation *in utero* [7].

Studies of epigenetic change over time have predominantly used cross-sectional approaches and have focused on adults [11-17] or on intrauterine development [18,19]. A small number of such studies have assessed age-associated DNA methylation across wider time-spans, encompassing childhood, adolescence and adulthood [20-23]. Consistently, age-associated changes in DNA methylation are more likely to involve (1) increases in methylation; (2) genes associated with development, signaling and regulation of transcription; and (3) regions involved in epigenetic reprogramming during embryonic

* Correspondence: jeff.craig@mcri.edu.au

† Contributed equally

²Early Life Epigenetics Group, MCRI, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia

Full list of author information is available at the end of the article

stem cell differentiation [12,14]. Since most of these studies have focused on CpG islands and promoter regions, age-associated epigenetic changes are incompletely characterized in relation to genomic coverage and life-course.

Longitudinal studies investigating aging and longevity have distinct advantages over cross-sectional designs, particularly in relation to controlling for genetic variation. For example, longitudinal studies directly query temporal sequences and pathways and individuals are studied rather than group averages. Longitudinal, array-based studies in blood from children in the first 1 to 5 years of postnatal life have shown similar results to adult studies with respect to gene function, genomic location and direction of age-related changes in DNA methylation, with the majority of age-related changes being observed in regions flanking CpG islands [24-26]. In adults, a high-resolution array-based study found a mixture of age-stable and age-dynamic variability throughout the methylome in adults [27]. Other studies of global DNA methylation in adults also showed a genetic influence on increase and decrease in DNA methylation [28] and a decrease in interspersed repeat DNA methylation over time [29].

A small number of epigenetic studies of ageing have focused specifically on twins. Cross-sectional studies have found that older monozygotic (MZ) twins differ more with respect to global and repetitive DNA methylation [30], a phenomenon referred to as 'epigenetic drift'. In contrast, using DNA from saliva, no evidence for such drift was found within MZ twins aged 21 to 55 years using a promoter/CpG island array [15]. A cross-sectional comparison of DNA methylation at the imprinted *IGF2/H19* locus in adolescent and middle-aged MZ twins also found no evidence of epigenetic drift within pairs [31]. To our knowledge, only one study has examined DNA methylation in twins longitudinally, measuring three loci in buccal DNA from 46 MZ pairs and 45 dizygotic (DZ) pairs at 5 and 10 years of age [8]. This study revealed (1) locus-specific variability in DNA methylation; (2) change over time in individuals; (3) within-pair correlation and (4) epigenetic drift. Taken together, these studies show that epigenetic change over time is likely to be regulated by many factors, potentially in a tissue-specific and genome context-dependent manner. Longitudinal epigenetic studies in twins offer tremendous potential to further our understanding of the relationship between genetics and other factors that specify inter-individual temporal change in DNA methylation profile in humans.

We have used the Infinium HumanMethylation450 BeadChip (HM450) platform, which interrogates >485,000 CpG dinucleotides and contains probes from CpG islands, shores (2 kb regions flanking CpG islands), shelves (2 kb regions flanking shores), sites from 1,500 bp upstream of transcription start sites through to gene bodies and

3' UTRs, in addition to intergenic regions, regions involved in epigenetic reprogramming during embryonic stem cell differentiation and enhancers [32,33]. Although repeats are not covered by these arrays and intergenic regions are not covered to the same depth as genic regions, the platform represents a significant step towards genome-scale coverage. Using the Infinium HM450 platform, we have performed a longitudinal study of DNA methylation at birth and age 18 months in DNA from buccal swabs from 10 MZ and 5 DZ twin pairs from the Peri/postnatal Epigenetic Twins Study (PETS) cohort [34]. We report a large degree of epigenetic change during the first 18 months of postnatal life, with strong regional genomic biases for rate of change over time. We also present evidence for pair-specific levels of epigenetic change, suggesting a complex interplay between environment, non-shared environment and stochastic factors in molding the early postnatal epigenome.

Results

Data pre-processing

Our initial analysis of HM450 data included normalization of previously identified differences between Infinium I and Infinium II probes [35] using the SWAN method [36]. Stringent quality control steps to assess probe performance (see Materials and methods) and removal of all probes on X and Y chromosomes to minimize sex-specific effects, resulted in 53/60 samples (Table 1) with data from 330,168 probes remaining for downstream analysis.

Determination of technical versus biological variation

First to assess the sensitivity to detect biological variability between co-twins versus technical variation, we performed replicate hybridizations of three MZ twin pairs both at birth and 18 months. We compared the level of variation between co-twins (biological variation) to the level of variation between each technical replicate sample (technical variation). Biological variation (twin 1 versus twin 2) consistently exceeded technical variation (twin 1 versus twin 1; twin 2 versus twin 2) for each twin pair (Figure 1). We determined the average level of differential methylation between all biological and technical replicate arrays using a moderated paired *t*-test with false discovery rate correction. Precisely 230,340/330,155 probes were differentially methylated (adjusted *P*-value <0.05) across all biological replicates, whilst 858/330,155 probes were found to vary (adjusted *P*-value <0.05) across all technical replicates of twin pairs.

Determining relationships between samples

Unsupervised hierarchical clustering of the entire dataset (Figure S1A in Additional file 1) revealed that most samples cluster according to age. The majority of co-twins also cluster together: 7/9 (78%) MZ co-twins cluster at

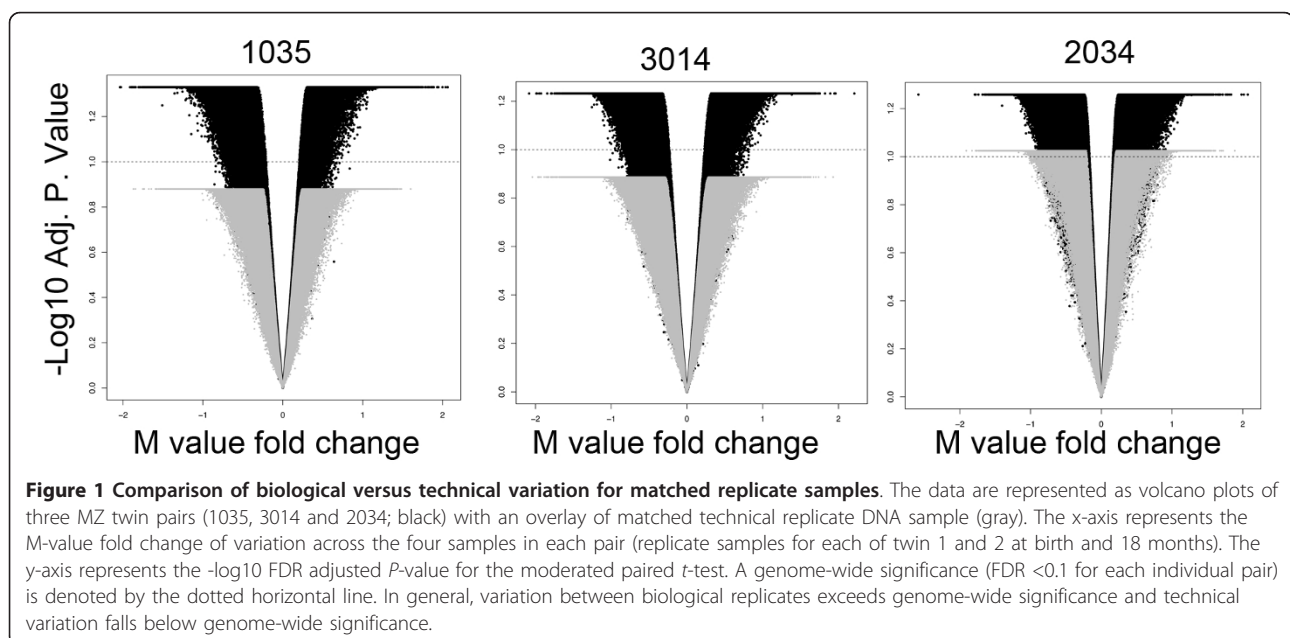
Table 1 Twin pair characteristics

Twin pair ID number	Zygosity ^a	Chorionicity ^b	Twin 1 sex	Twin 2 sex	Gestational age	Birth weight discordance (%) ^c	Samples removed after QC ^d
1016	MZ	DC	M	M	37	5.4	T2_18
1022	MZ	MC	M	M	38	12.4	T1_18, T2_18
1024	MZ	DC	F	F	37	12.1	T2_B
1032	DZ	DC	F	F	37	43.3	T1_18
1035 ^e	MZ	MC	F	F	35	30.8	
1042	DZ	DC	M	M	30	19.8	
1046	MZ	DC	M	M	37	6.2	
1057	DZ	DC	M	M	37	13.3	
1058	MZ	DC	M	M	36	22.0	
1072	DZ	DC	F	F	37	14.5	
1107	MZ	MC	F	F	33	8.1	T1_18
1126	MZ	MC	F	F	32	27.3	T2_18
2034	MZ	DC	F	F	36	3.6	
3006	DZ	DC	M	M	37	3.6	
3014	MZ	MC	M	M	36	18.0	

^aMZ, monozygotic; DZ, dizygotic. ^bMC, monochorionic; DC, dichorionic. ^c $[(\text{Weight of heaviest twin} - \text{Weight of lightest twin}) / \text{Weight of heaviest twin}] \times 100$. ^dT1/T2, twin 1/2; B, birth sample; 18, 18 months sample. ^eTwin-to-twin transfusion syndrome.

birth and 6/6 (100%) at 18 months, while with DZ co-twins, 4/5 (80%) cluster at birth and 2/4 (50%) cluster at 18 months. To explore the variation in this dataset attributable to the effect of sequence variation on methylation values via *cis* genetic effects or probe hybridization, we performed hierarchical clustering selectively for probes overlapping known SNPs as defined by the HM450 SNP manifest (version 3, 103,148 probes). The results compared well to the full dataset: 7/9 MZ co-twins cluster at birth and 6/6 cluster at 18 months, while 4/5 DZ co-twins cluster at birth and 3/4 cluster at 18 months (Figure S1B

in Additional file 1). Restricting this analysis to probes with reported SNPs at the CpG site assayed by the probe (2,527 probes in this data set) resulted in 8/9 MZ co-twins clustering at birth and 6/6 at 18 months (Figure S1C in Additional file 1). Interestingly, on average for this set of probes, DZ twins did not cluster with their co-twin; rather, DZ twins at birth clustered with their matched samples at 18 months. Thus, data for such probes are likely to reflect the genotype of the individual rather than representing purely methylation levels. A random sampling of the same number of SNP-associated probes did not reproduce this



clustering (data not shown), indicating this effect did not represent a sampling bias. These results suggest that SNP-containing probes account for little variation in the overall data set, with the exception of probes with SNPs at the CpG site assayed.

Identification of age-associated differentially methylated probes

To identify specific sites of differential DNA methylation associated with age, we used an empirical Bayes method [37] to compare birth samples with matched 18-month samples in all individuals and performed a probe-wise moderated paired *t*-test for differential methylation. Using this approach we found that 30.1% (99,198) probes changed significantly over time (adjusted *P*-value <0.05). These age-associated differentially methylated probes (aDMPs) changed by a mean β of 0.031 (3.1%) per year. Adding a further stringent cutoff of >20% absolute change over time to minimize technical effects [38] resulted in 0.8% (2,632) probes classified as stringent aDMPs (Table S1 in Additional file 2). Of these aDMPs, 87% showed a gain in DNA methylation over time whereas 13% showed reduced methylation (Figure 2a). We selected candidate aDMPs for validation based on their ranked change in methylation β value from birth to 18 months. The Sequenom MassArray EpiTyper platform was used to provide an independent measure of DNA methylation at aDMPs and confirmed the validity of the HM450 dataset. Using this approach, we

confirmed that aDMPs identified by HM450 analysis are also representative of methylation at surrounding CpG sites (Figure 2b). Ontology and pathway analyses of the aDMP-associated genes showed an over-representation of cell development, morphogenesis (especially neuronal cells), and GTPase signaling pathways (Table 2; Tables S2 and S3 in Additional file 2). In order to determine whether aDMPs were more likely to occur at specific regions in the genome, we calculated the observed/expected frequency (enrichment) of genomic locations annotated in the HM450 manifest and assigned *P*-values with hypergeometric means tests. Intergenic regions were most likely to show changes in DNA methylation from birth to 18 months (Figure 3, grey bars; enrichment = 6.0 \times), followed by enhancers (2.5 \times) and 'open sea' regions >4 kb distant from CpG islands (1.7 \times). Promoters and CpG islands, but not their flanking shores and shelves, were less likely to show changed methylation over time (Figure 3; relative enrichment of 0.23 \times and 0.39 \times , respectively

Identification of age-associated differentially methylated regions

In order to identify larger regions of coordinated methylation change over time, we adopted a recently published differentially methylated region (DMR)-finding method [39]. This 'bump hunting' method identifies genomic regions in which clusters of consecutive CpG sites exhibit change over time in the same direction. Estimates were

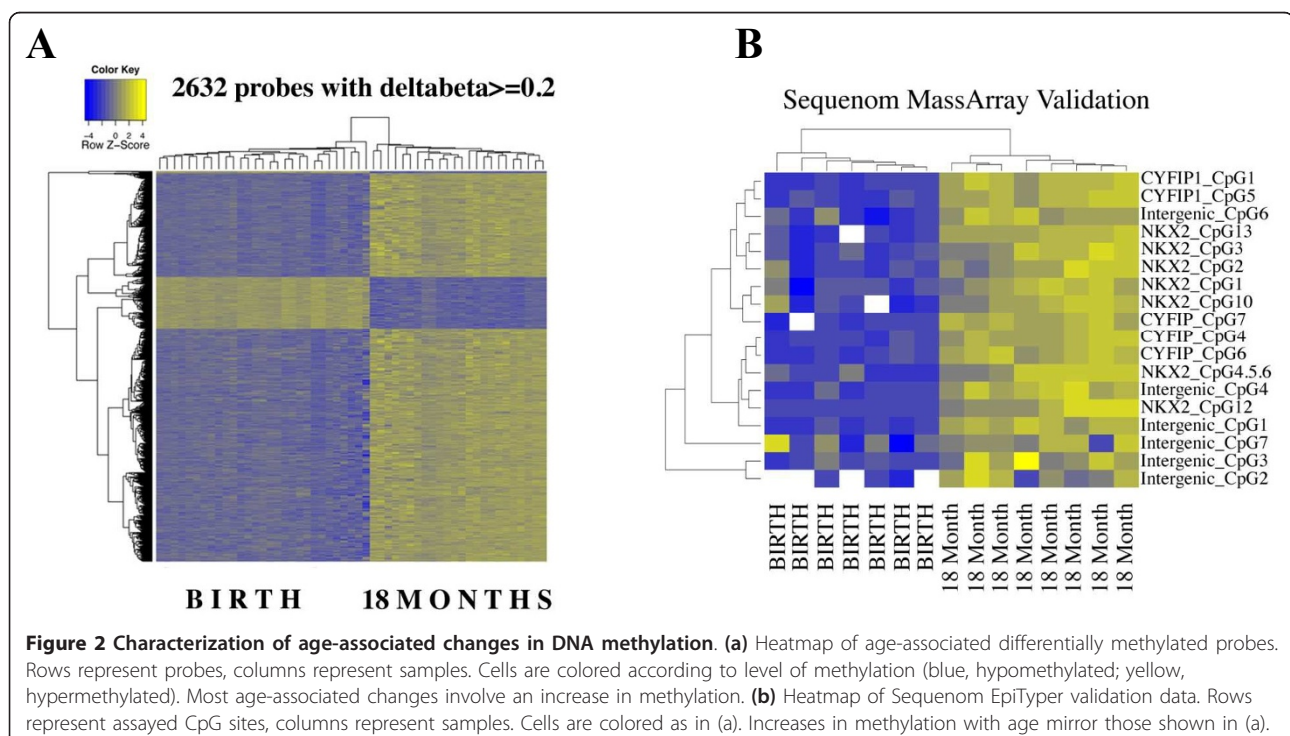


Table 2 Ontology enrichment analysis of age-associated differentially methylated probes

GO term	Description	Adjusted P-value
GO:0000904	Cell morphogenesis involved in differentiation	0.0035
GO:0051056	Regulation of small GTPase-mediated signal transduction	0.0042
GO:0046578	Regulation of Ras protein signal transduction	0.0075
GO:0030182	Neuron differentiation	0.0157
GO:0031175	Neuron projection development	0.0192
GO:0048667	Cell morphogenesis involved in neuron development	0.022
GO: 0007409	Axonogenesis	0.026
GO:0048812	Neuron projection morphogenesis	0.027
GO:0000902	Cell morphogenesis	0.038

Gene Ontology (GO) terms significant at an adjusted *P*-value <0.05 [81] are shown. For full analysis, see Table S2 in Additional file 2.

obtained for aDMRs by computing group medians and obtaining a value for the smoothed estimate that exceeds a *t*-statistic cutoff of 0.995. Using these criteria, we defined 897 aDMRs consisting of 4 or more consecutive probes changing in methylation between birth and 18 months (aDMPs). These aDMRs ranged in size from 33 to 1,698 bp. Twelve of these regions contained ten or more consecutive probes within approximately 1 kb of

each other (Table S4 in Additional file 2 with an example shown in Figure 4). Of all aDMRs, 44% are located within 5 kb of a transcriptional start site, compared to 29% for aDMPs (Figure S2 in Additional file 1 and Table S4 in Additional file 2). Ontology analysis indicated that the aDMRs were significantly enriched for biological processes associated with cellular and organ development and in DNA binding (FDR <0.05; Table 3). As the sixth largest

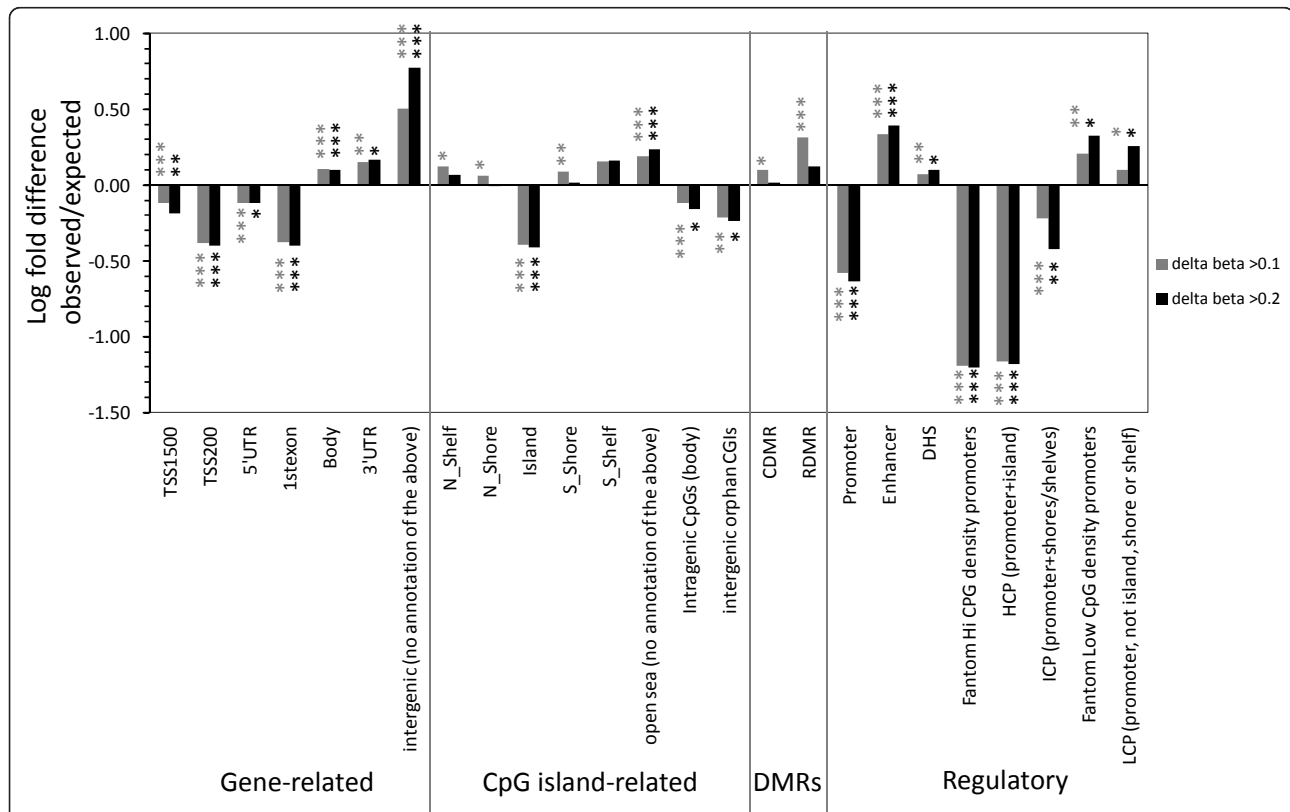


Figure 3 Enrichment of aDMPs by genomic location. Log-fold difference of enrichment (observed/expected frequency) in the aDMPs with *P* < 0.05 and delta beta > 0.1 (*n* = 14,629) and > 0.2 (*n* = 2,632) for specific genomic locations, grouped by association with genes, CpG islands, known DMRs and regulatory regions. Positive values indicate enrichment and negative values indicate depletion in the aDMP dataset. *P*-values: **P* < 0.05; ***P* < 1 × 10⁻²⁰, ****P* < 1 × 10⁻⁵⁰.

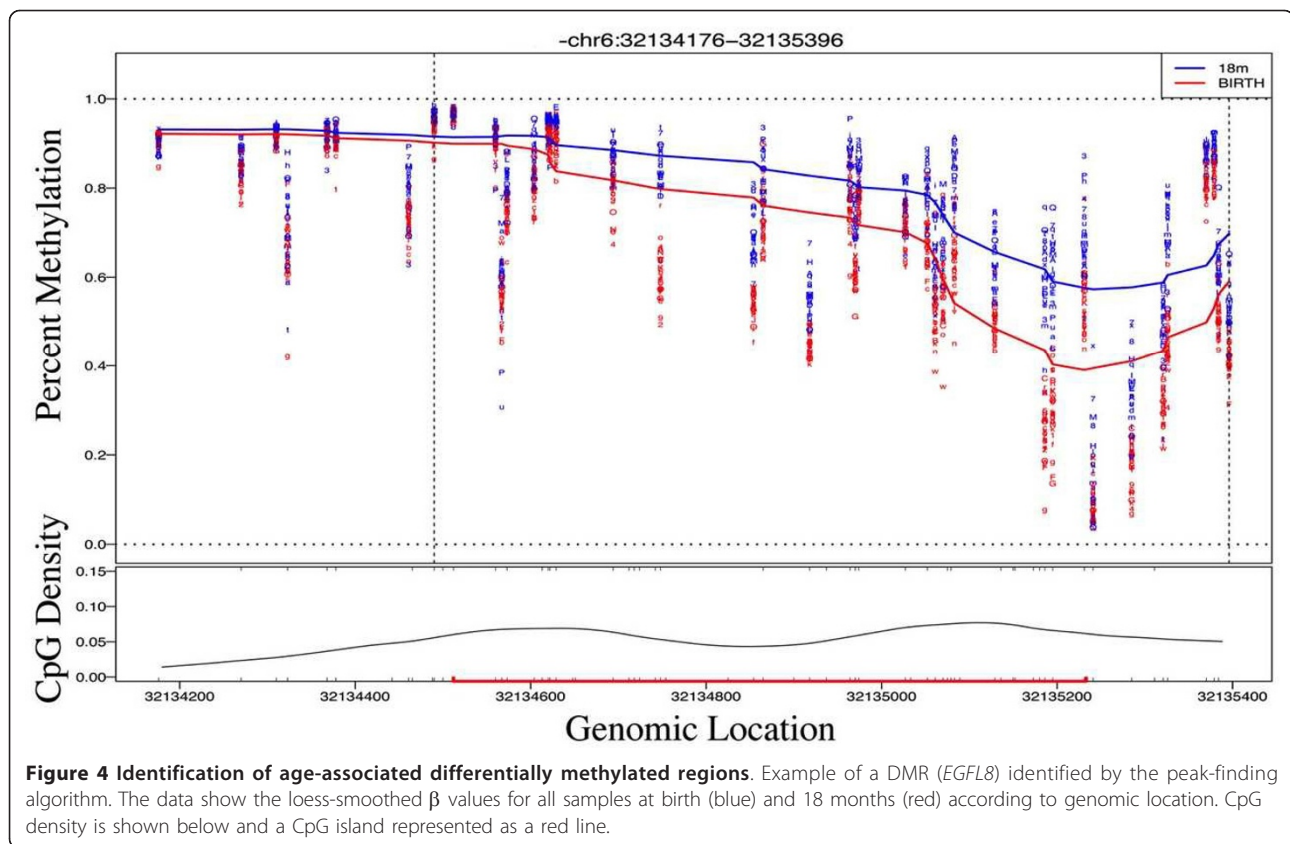


Table 3 Ontology enrichment analysis for age-associated differentially methylated regions

GO term	Description	$-\log_{10}$ binomial <i>P</i> -value
GO:0048513	Organ development	17.68
GO:00048869	Cellular development process	15.59
GO:0030154	Cell differentiation	15.33
GO:0010033	Response to organic substance	14.47
GO:0009887	Organ morphogenesis	14.00
GO:0006359	Regulation of transcription from RNAs pol III promoter	9.66
GO:0000975	Regulatory region DNA binding	8.73
GO:0044212	Transcription regulatory region DNA binding	8.00
GO:0045945	Positive regulation of transcription from RNA pol III promoter	7.51
GO:0048598	Embryonic morphogenesis	6.80
GO:0016480	Negative regulation of transcription from RNA pol III promoter	6.67
GO:0003205	Cardiac chamber development	5.71
GO:0030326	Embryonic limb morphogenesis	4.4
GO:0060173	Limb development	3.82
GO:0035108	Limb morphogenesis	3.62
GO:0048546	Digestive tract morphogenesis	3.10

Gene Ontology (GO) terms significant at an adjusted *P*-value <0.05 [81] are shown.

DMR (Table S4 in Additional file 2) and a representative of a number of the top age-associated ontologies, DNA methylation was validated at the cytoplasmic FMR1 interacting protein 1 (*CYFIP1*) gene in all samples using the Sequenom MassArray EpiTyper platform (Figure 2b).

Epigenetic discordance within twin pairs at birth and 18 months

Within-pair epigenetic discordance resulting from non-shared environmental factors has been postulated to underscore variation in phenotypic traits [40,41]. We

examined discordance in DNA methylation profile within twin pairs at birth and at 18 months of age. We calculated twin discordance as the absolute difference in β methylation values within pairs in birth samples and separately for 18-month samples. We ranked all probes according to average within-pair discordance at each age and performed 'ranked-list' ontology, which differs from 'gene-list' ontology in that there is no requirement for a predefined cutoff. All probes on the array were ranked by their scores for average within-pair discordance at each age (most discordant to least discordant), and the ranked list of probes was analyzed by the GOrilla bioinformatics tool [42] to identify ontology terms over-represented at the top of the list, compared with the bottom. We found that the most discordant genes at birth were consistently enriched for ontology terms associated with RNA metabolism, including spliceosome components and transcription factors (Table S5 in Additional file 2). At 18 months of age, the most discordant genes were associated with a similar set of gene ontologies as seen at birth (Table S6 in Additional file 2). The genes with discordant probes at both time points include a wide array of spliceosome components (for example, *WDR83* and *CWC22*), zinc finger proteins (for example, *ZNF267*, *ZBTB1*, *ZNF10*), ribosomal proteins (for example, *RPS26*, *RPL15*, *RPL12*) and transcription factors (for example, *MAMLL1*, *HOXB13*).

We next investigated the distribution of DNA methylation discordance across genomic regions to determine whether discordance is more likely to occur at specific genomic locations. We have shown previously, using HM27 arrays, that median within-pair methylation discordance increased with increasing distance from CpG islands in three tissues (cord blood mononuclear cells, human umbilical vein endothelial cells and placenta) in both MZ and DZ twins at birth [7]. As the HM27 array focuses primarily on gene promoters and CpG islands, we repeated this analysis taking advantage of the diversity of genomic locations contained within the HM450 arrays. We calculated absolute within-pair discordance as before, and plotted probe discordance across genomic location at birth and at 18 months. The distribution of discordance values was consistent across all genomic annotations targeted on the array, with no evidence of regional enrichment (Figure S3A in Additional file 1). Similar results were observed selecting the top 10,000 most variable probes, or alternatively when the analysis was performed separately at birth and 18 months separately for both MZ and DZ twins (data not shown). We then filtered the dataset to include only probes present on the HM27 arrays and found evidence of higher levels of discordance around shores and shelves of CpG islands (Figure S3B in Additional file 1), which is consistent with our previously published observation with this platform [7].

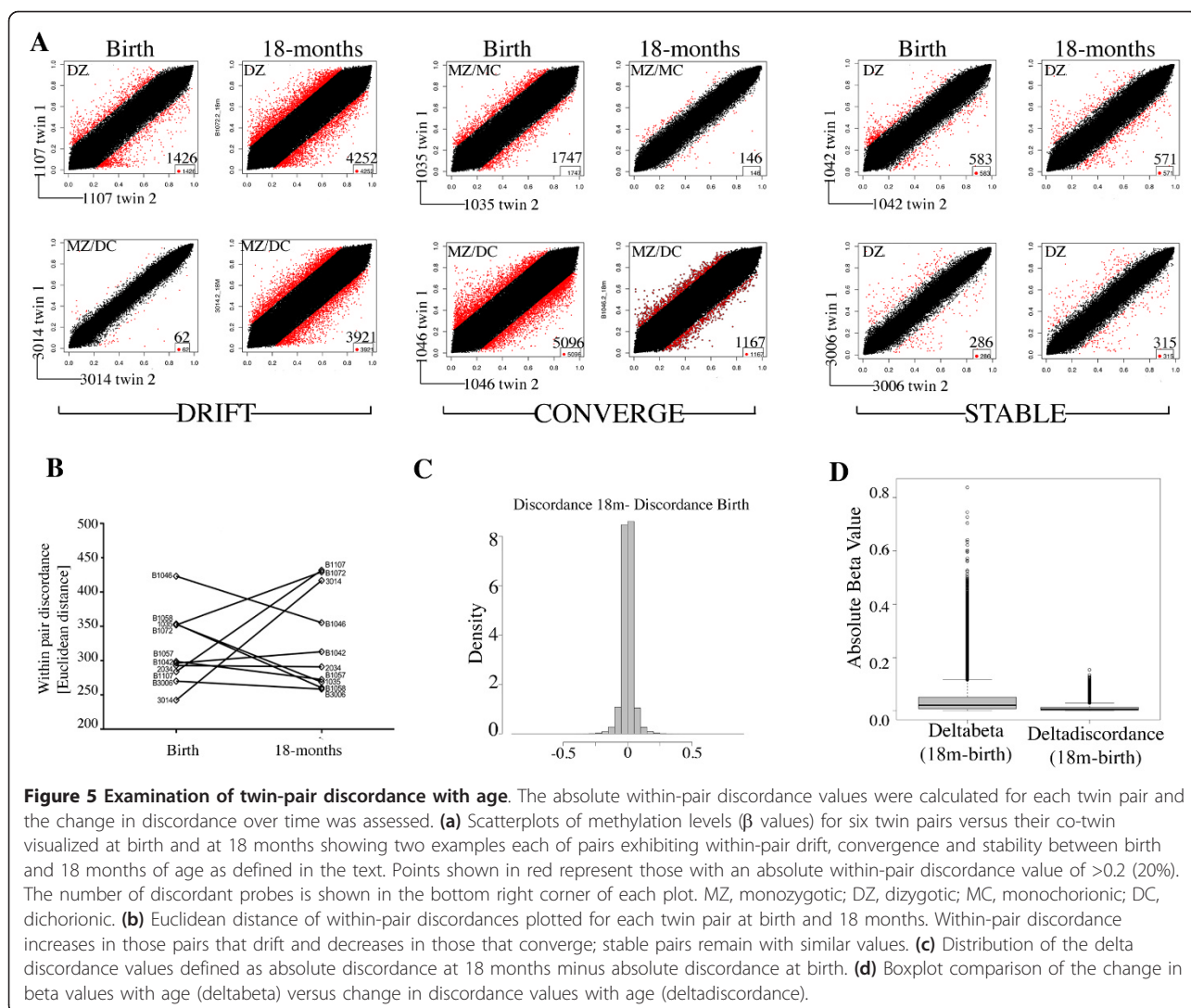
Level of change in epigenetic discordance (drift versus convergence) over the first 18 months is a pair-specific phenomenon

Since previous cross-sectional studies suggest that epigenetic discordance in twins increases with age [30], we next investigated the degree of epigenetic drift from birth to 18 months of age within our twin pairs. The probe-wise level of within-pair discordance for CpG sites exhibiting a β -discordance value of greater than 0.2 (>20% discordant) was visualized at each age on scatterplots (Figure 5a, points in red). In contrast to the anticipated drift associated with age, we observed that the degree of within-pair discordance over time varies in a pair-specific manner (Figure 5a), with some pairs becoming more discordant in 18 month samples compared to birth samples (that is, epigenetic drift), some pairs becoming less discordant in 18 month samples, which we termed 'convergence', and others similarly discordant at both ages ('stable'). This was supported by Euclidean distance measures of twin discordance [7] (Figure 5b). These phenomena were not associated with zygosity or chorionicity, nor influenced by the effects of probes targeting SNPs on the array (Figure S4 in Additional file 1). We further calculated the change in discordance with age (delta discordance), as the difference in twin discordance (absolute values) from birth to 18 months. The distribution of delta discordance values was strongly centered about zero, with no evidence for overall skewing with age (Figure 5c). A comparison of the magnitude of the absolute values of differences in within-pair discordance over time (delta discordance) compared to the absolute values for methylation change over time (18 months - birth) indicated that age-related changes are far greater on average than changes to within-pair discordance (Figure 5d).

We next sought to explore epigenetic drift and convergence in more detail. As there were no probes that showed consistent differences across all pairs, our aim was to determine whether we could identify any gene ontologies associated with probes that consistently 'drift', 'converge' or remain stable among our twins over time. To address this we grouped our twin pairs according to their observed temporal methylation discordance patterns ('drifting' or 'converging'), measured as values of change in discordance (delta discordance). Within both of these categories, we calculated the within pair delta discordance for each probe, and ranked all probes on the average delta discordance across pairs. We performed gene ontology analysis on the ranked lists for each 'drifting' and 'converging' category and found both were similarly enriched for genes involved in development and morphogenesis (Table 4).

Discussion

This study, examining DNA methylation profiles in buccal cells of young twins, has further confirmed the



highly dynamic nature of the human epigenome postnatally, in agreement with previous studies in other tissues [20,21,24-26,43]. Buccal cells represent a key bio-resource for age- and disease-associated epigenetic association studies [8,12,44,45]. From a previous study [46] and our own unpublished data, this sample type comprises $>90\%$ squamous epithelial cells with $<10\%$ blood cells. Moreover, we minimized cell heterogeneity due to immune reactions by not collecting from infants with mouth infections. Almost a third of all HM450 probes in our final dataset showed significantly changed DNA methylation levels at $FDR < 0.05$. Furthermore, the average absolute methylation change in these aDMPs was 4.7% (3.2% per year over 18 months) and almost 3% of aDMPs exhibited an absolute methylation change of $>20\%$. These changes are similar in magnitude to those seen in blood from birth to one year of age using HM27 arrays and $FDR < 0.05$ (3.9% of probes with changes $>20\%$; average change of 9.2% per year) [24]

and in T cells from birth to one year of age using HM450 arrays and $FDR < 0.01$ (2.7% of probes with changes $>20\%$; average change of 14% per year) [25]. In addition, our findings are of similar magnitude to a cross-sectional study of DNA methylation in the prefrontal cortex from human cadavers using HM27 arrays [47]. In combination with other cross-sectional studies [20,22], these cumulative data support the idea that rate of change of DNA methylation in the genome in any one tissue is highest *in utero*, possibly reflecting extensive cellular differentiation during organogenesis, and then declines in childhood, with a further drop in adulthood [20,21,43].

In the current study, most (approximately 90%) age-associated changes involved an increase in DNA methylation over time. This agrees with previous longitudinal studies of early childhood [24-26] and cross-sectional studies of placenta throughout gestation [48], peripheral blood in children [20], peripheral blood in adults

Table 4 Ontology enrichment analysis for drifting and converging pairs

GO term	Description	Adjusted P-value
Drifting pairs		
GO:0032502	Developmental process	3.21E-12
GO:0048856	Anatomical structure development	8.34E-10
GO:0048869	Cellular developmental process	2.15E-09
GO:0048598	Embryonic morphogenesis	2.61E-07
GO:0009653	Anatomical structure morphogenesis	9.50E-07
GO:0007389	Pattern specification process	2.36E-06
GO:0050793	Regulation of developmental process	3.86E-06
GO:0007166	Cell surface receptor signaling pathway	5.33E-06
GO:2000026	Regulation of multicellular organismal development	1.06E-05
GO:0023051	Regulation of signaling	2.59E-05
Converging pairs		
GO:0048856	Anatomical structure development	7.43E-08
GO:0032502	Developmental process	1.01E-07
GO:0048598	Embryonic morphogenesis	2.14E-05
GO:0050793	Regulation of developmental process	2.99E-05
GO:0022610	Biological adhesion	3.66E-05
GO:0007155	Cell adhesion	3.05E-05
GO:0045595	Regulation of cell differentiation	7.62E-05
GO:0048869	Cellular developmental process	1.95E-04
GO:0023051	Regulation of signaling	3.07E-04
GO:2000026	Regulation of multicellular organismal development	3.66E-04

The top ten Gene Ontology (GO) terms ranked by FDR (q values) ranked by mean delta discordance across 'drifting' and 'converging' pairs.

[12,13,16], buccal cells in adults [12] and in a comparison of differences in various tissues between fetal and adult tissues [22]. A recent study comparing the entire methylomes of a newborn and a centenarian using bisulfite sequencing [23] observed an age-associated decrease in methylation in most genomic regions, including interspersed repetitive DNA, intergenic and intragenic regions. Although additional studies are needed to reconcile these discrepancies, it is also likely that age-associated methylation changes are dependent on genomic and tissue context, on the methylation analysis platform used and on sample size.

We found that aDMRs are more likely to be in intergenic and intragenic regions (Figure 3). Such regions were also enriched in aDMRs identified in multiple tissues in rats [49] and mice [50]. The intergenic regions identified in our study are single copy and overlap with enhancers (Figure 3). Such regions undergo the most dynamic changes accompanying differentiation of pluripotent stem cells [51]. Almost a third of aDMPs identified, and almost a half of aDMRs, lay within 5 kb of transcriptional start sites (Figure S2 in Additional file 1), implicating such regions in the regulation of gene expression. The higher proportion of aDMRs (44%) compared to aDMPs (29%) around gene promoters most likely reflects the higher CpG density and co-methylation (locally correlated methylation) within these regions [52].

However, this could also be due at least in part to the relatively wider HM450 probe spacing in intergenic regions.

Ontology and pathways analysis showed that approximately three quarters of the genes associated with stringent aDMPs and all aDMRs are implicated in development and morphogenesis (Tables 2 and 3; Table S2 in Additional file 2). Signaling pathways, including those based on GTPase signal transduction, pathways intrinsic to development, were also enriched, as has been observed in other studies of age-associated changes in methylation [50,53]. A bias towards genes involved in development has been seen in cross-sectional [13,20,54,55] and longitudinal [25,27] studies of human aging and in a study of embryonic stem cell differentiation [53].

It is interesting to note that approximately half of the top ten aDMP ontologies related to neural development (Table 2), a finding shared with previous studies of methylation in saliva [15] and blood cell fractions and buccal epithelium from adults [12] and in a large meta-analysis of multiple HM27 datasets from human brain and blood [55]. Further studies are needed to ascertain whether these findings relate to biases related to the large proportion of the genome expressed in the brain.

We found that despite a trend towards increased methylation with age in all regions of the genome, CpG-dense promoter regions were particularly depleted in aDMRs

and CpG-poor promoters were moderately increased (Figure 3), contrary to previous cross-sectional [13,15,16,56] and longitudinal [26] studies of ageing-associated methylation change. This disagreement is most likely because HM27 arrays are enriched in CpG-dense CpG islands. However, our data agree with findings from studies using methods that include intergenic regions, that low CpG density promoters are enriched in mouse aDMRs [50], during differentiation of human embryonic stem cells [57] and between birth and very old age [23].

Epigenetic discordance at birth and 18 months of age

Very few genome-wide studies of methylation or expression have been performed on buccal cells. One such study, of buccal cells collected from 20 twins aged 13 to 14 years using a low resolution CpG island array, found no significant methylation differences within pairs [44]. However, a study of smoking-induced differential gene expression in buccal cells identified a differentially expressed network of genes with, at the hub, transcription factors REL and CREB [58], which are among the top 10% most discordant genes at birth and 18 months in our data (Tables S5 and S6 in Additional file 2). Despite the extensive longitudinal changes in DNA methylation described above, we found that, in general, probes located within genes associated with RNA metabolism (for example, spliceosome components) and control of gene expression (for example, transcription factors) were consistently more discordant within twin pairs at both birth and 18 months of age. Of interest, this class of genes has previously been shown to have altered levels of transcription in buccal cancer [59].

Epigenetic drift and convergence

In the current study, we found that a summed value (Euclidean distance) of epigenetic discordance across hundreds of thousands of loci can vary between and within pairs and can increase or decrease over time. In accordance with our genome-scale findings, a longitudinal study of DNA methylation at seven imprinted gene loci in buccal cells between birth and one year of age showed that inter-individual variation similarly increased, decreased or remained similar in singletons and that the direction of change could differ between individuals [45]. A longitudinal study of DNA methylation at three genes in buccal cells in 46 MZ and 45 DZ twin pairs found that methylation drifted in some pairs and converged in others over time [8]. Similar results were found for MZ and DZ twins and a role for genetic, shared and non-shared environmental factors, dependent on genomic location, in these longitudinal changes was postulated [8]. For MZ pairs, changes in within-pair discordance must be influenced solely by stochastic and non-shared environmental factors. Evidence for the latter comes

from our previous studies of methylation in newborn twins [6,7,60] and from a cross-sectional study of DNA methylation in seven genes in whole blood from >200 MZ twin pairs aged 18 to 89 years [61]. Data from a longitudinal, genome-scale study of DNA methylation (using HM450 arrays) in whole blood from an independent cohort of young adults (aged 22 to 32 years) also provides evidence of genome-scale methylation drift and convergence defined by changes in Euclidean distance over time (Figure S5 in Additional file 1).

Epigenetic drift has been postulated to arise from the cumulative effects of (non-shared) environment and stochastic events [30,62,63], the latter influenced by epigenetic events such as promoter occupancy by transcription factors [64] and by errors made during the maintenance of DNA methylation profile following DNA replication [30,63]. Recent studies suggest that epigenetic drift may also reflect differing rates of change of methylation among the population [65]. Furthermore, others have argued that epigenetic variability (or noise) is itself genetically programmed and has evolved to mediate some degree of plasticity (via canalization) [66]. In contrast, we suggest that 'convergence' may involve sites of methylation equalization between co-twins, possibly reflecting regression to the mean as a contributing factor. Regression to the mean is a phenomenon in which it is a statistical certainty that individual phenotypes, such as growth patterns [67], shift to the population mean over time [68]. This explains why twins with birth weight discordance become more similar over time [69] and can be understood in terms of twin-specific uterine-specific restrictions being replaced postnatally by a greater degree of shared environment [69-71]. Indeed, the twins in the current study had a median weight discordance [(Weight of the heavier twin - Weight of the lighter twin)/Weight of the heavier twin] of 13.3% at birth and 2.8% at 18 months. Although caution is needed with interpretations from a small sample size, we note that 'converging' pairs were more likely to start with a higher within-pair discordance (mean Euclidean distance = 375) than the drifting pairs (mean Euclidean distance = 295) (Figure 5b), although this difference did not reach significance ($P = 0.11$). Clearly, larger longitudinal twin-based studies are needed to further investigate factors contributing to epigenetic drifting and convergence over time.

Conclusions

We have conducted the first longitudinal study of epigenetic change in buccal cells in twins from birth, using a validated, genome-scale methylation array. We have shown evidence that the epigenetic profile of both MZ and DZ twin pairs can exhibit epigenetic drift or convergence early in postnatal development. As genes involved in development exhibited the largest absolute changes in methylation over time and the largest, smaller-scale

changes within twin pairs, we conclude that the epigenetically driven developmental program is influenced to some extent by stochastic and/or non-shared environmental factors. Thus, canalization may be influenced by such factors, in addition to genetic factors as suggested by Waddington [72,73].

Materials and methods

Subjects, tissues and DNA extraction

Sample collection from twins at the time of delivery was carried out with appropriate human ethics approval from the Royal Women's Hospital (project number 06/21), Mercy Hospital for Women (project number R06/30), and Monash Medical Centre (project number 06117C), Melbourne and the study was conducted according to the Declaration of Helsinki principles. The twin pairs chosen for methylation array analysis are shown in Table 1. The 10 MZ pairs and 5 DZ pairs shared a similar sex ratio, gestational age and birth weight to the full group of 250 pairs. Buccal cells were collected with Catch-all Sample Collection Swabs (EPICENTRE Biotechnologies, Madison, WI, USA) and were stored at -20°C until DNA extraction, which was performed as previously described [60].

Infinium HumanMethylation450 BeadChip data acquisition and processing

DNA samples (1 μg) were bisulfite converted using the Methyl EasyXceed bisulfite modification kit (Human Genetic Signatures, North Ryde, Australia), according to the manufacturer's instructions. Conversion efficiency was assessed by bisulfite-specific PCR. DNA samples were hybridized to Illumina Infinium Human Methylation450 (HM450) BeadChip arrays according to the manufacturer's instructions. Raw intensity data (IDAT) files were imported into the R environment (version 2.14.1) [74] and processed using the *minfi* package [75]. All analyses were performed in R using packages available from the Bioconductor project [76]. Data quality was assessed in *minfi* using plots derived from various control probes on the array. Poor performing probes defined as those with an average detection P -value >0.001 in one or more samples were removed from the analysis ($n = 132,113$). Data from five samples with an average detection P -value >0.05 and with evidence of poor bisulfite conversion efficiency were removed completely. Probes on the X and Y chromosomes were also discarded from all samples. The resulting data were pre-processed using the Illumina method within *minfi* and subset-quantile within-array normalization was performed [36] for combined normalization of Infinium type I and type II probes. The log₂ ratio of methylated probe intensity to unmethylated probe intensity were calculated in *minfi* and the resulting M-values [77,78] were quantile normalized between arrays using the *limma* package [79]. Sample quality was

further assessed using hierarchical clustering plots available in *minfi* and *lumi* [77] packages. Following this, three additional samples were removed as outliers constituting a final data set of 330,168 probes and 53 samples.

Statistical analysis

Exploratory analysis of sample relationships was performed using unsupervised hierarchical clustering analysis with the Euclidean distance and complete linkage algorithm, and dendrogram was created using *gplots* [80]. Differential methylation analysis was performed on M-values using the *limma* package using a cutoff of FDR-corrected P -values <0.05 [81] and delta beta values >0.2 . To study discordance among co-twins at the probe-level, a linear model was fitted to the M-values with twin-pair as a predictive factor to model the twin relationship. The level of discordance among co-twins was interpreted as the residual measurement for each CpG from the model-fit. For enrichment analysis, gene sets were populated with probe IDs using the annotated regions provided in the Illumina HM450 manifest file (version 1.1). Annotations used were classified as gene-related (TSS1500 and TSS200, regions from -1500 to -200 and -200 to the transcriptional start site respectively, 5' UTRs, first exons, gene bodies, 3' UTR and intergenic (no gene annotation)); CpG island-related (islands (also split into intragenic and intergenic)), shores (0 to 2 kb flanking islands), shelves (2 to 4 kb flanking islands) and open sea (>4 kb from islands) [82]; DMRs (associated with cancer (CDMRs) and induced pluripotent stem cell reprogramming (RDMRs); [83] and regulatory regions (promoters, enhancers and DNase hypersensitivity sites, likely to be a mixture of promoters and enhancers [84,85]). Boxplots were produced to graph each category by discordance score. The 'bump-hunting' methods described by Jaffe and colleagues [39] were implemented using the *charm* package available in Bioconductor [86]. We used the 'dmrFinder' algorithm without covariate adjustment, using the default SPAN settings and specifying a minimum four probes, and a t-statistic cutoff to identify probes as being in a DMR at 0.995. For gene ontologies the GOrilla bioinformatics tool [42] was used to perform ranked-list ontology using the entire array content ranked by scores for discordance. Gene-list ontology enrichment was performed on significant gene lists (FDR <0.05) using the DAVID bioinformatics tool under the default settings [87]. Pathway analysis data were analyzed through the use of Ingenuity Pathway Analysis (Ingenuity Systems, Redwood City, CA, USA). The analysis tool GREAT (Genomic Regions of Annotations Tool) [88] was used to analyze the functional significance of aDMRs using the single nearest gene association rule within a 100 kb window.

Sequenom MassArray target validation

Target validation was performed using the Sequenom MassArray EpiTYPER (Sequenom, San Diego, CA, USA) performed as previously described [18,60]. Amplicons were designed using Sequenom EpiDesigner software. Primers are listed in Table S7 in Additional file 2. In brief, amplification was performed after bisulfite conversion of genomic DNA with the MethylEasy Xceed bisulfite conversion kit (Human Genetic Signatures, North Ryde, Australia). All PCR amplifications and downstream processing were carried out at least in duplicate and the mean methylation level at specific CpG sites determined. Raw data obtained from MassArray EpiTYPING were cleaned systematically using an R-script to remove samples that failed to generate data for more than 70% of CpG sites tested [60]. Also, technical replicates showing $\geq 10\%$ absolute difference from the median value of the technical replicates were removed and only samples with at least two successful technical replicates were analyzed. Samples were compared across each analyzable CpG site in the amplicon, as well as the mean across the whole amplicon.

Data availability

Array data described in this manuscript have been submitted to the Gene Expression Omnibus public repository and are freely available under the accession number GSE42700.

Additional material

Additional file 1: Supplementary figures and legends.

Additional file 2: Supplementary tables and legends.

Abbreviations

aDMP/R: age-associated differentially methylated probe/region; bp: base pair; DMR: differentially methylated region; DZ: dizygotic; FDR: false discovery rate; HM450: Infinium HumanMethylation 450 BeadChip array platform; MZ: monozygotic; RDMR: reprogramming differentially methylated probe/region; SNP: single nucleotide polymorphism; UTR: untranslated region.

Authors' contributions

RS and JMC conceived the study. YJL performed the Sequenom validation. DM, LG, MNC and YJL analyzed the data. MO contributed the Finnish methylation dataset. JMC and DM wrote the first draft of the manuscript. All authors participated in discussions related to analysis and interpretation, have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We wish to thank Ruth Morley, John Carlin, Clinical Epidemiology and Biostatistics Unit, MCRI, Mark Umstad, Royal Women's Hospital, Melbourne, Euan Wallace, Monash Medical Centre, Melbourne and Mark Permezel, Mercy Hospital For Women, Melbourne for their contributions to establishing the PETS cohort; Sarah Healy, Tina Vaiano, Nicole Brooks, Jennifer Foord, Sheila Holland, Anne Krastev, Siva Illancheran and Joanne Mockler for recruitment

and sample collection; Research Assistant Xin Li, Technical officer Anna Czajko, Study Coordinator Geraldine McLroy, and all mothers and twins that participated in this study. We also thank Jaakko Kaprio, Khadeeja Ismail and Kirsi Pietiläinen for sharing unpublished data and Ben Ong for help with Sequenom analysis. This work was supported by grants from the Australian National Health and Medical Research Council (grant numbers 437015 and 607358 to JMC and RS), the Financial Markets Foundation for Children (grant number 032-2007) and by the Victorian Government's Operational Infrastructure Support Program. RS is supported by a NHMRC Senior Research Fellowship. JMC and DM would also like to acknowledge financial support from the Murdoch Childrens Research Institute.

Author details

¹Cancer, Disease and Developmental Epigenetics, Murdoch Childrens Research Institute (MCRI), Flemington Road, Parkville, Victoria 3052, Australia. ²Early Life Epigenetics Group, MCRI, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. ³Department of Paediatrics University of Melbourne, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. ⁴Bioinformatics Unit, MCRI, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. ⁵Hjelt Institute, Department of Public Health, University of Helsinki, Mannerheimintie 172, 00300 Helsinki, Finland. ⁶Current address: Division of Leukaemia and Cancer Research, Telethon Institute for Child Health Research, Centre for Child Health, University of Western Australia, 100 Roberts Road, Subiaco, Western Australia 6008, Australia.

Received: 20 December 2012 Revised: 24 April 2013

Accepted: 22 May 2013 Published: 22 May 2013

References

1. Feinberg AP: Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007, **447**:433-440.
2. Foley DL, Craig JM, Morley R, Olsson CA, Dwyer T, Smith K, Saffery R: Prospects for epigenetic epidemiology. *Am J Epidemiol* 2009, **169**:389-400.
3. Petronis A: Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 2010, **465**:721-727.
4. Bell JT, Spector TD: A twin approach to unraveling epigenetics. *Trends Genet* 2011, **27**:116-125.
5. Bell JT, Saffery R: The value of twins in epigenetic epidemiology. *Int J Epidemiol* 2012, **41**:140-150.
6. Gordon L, Joo EJ, Andronikos R, Ollikainen M, Wallace EM, Umstad MP, Permezel M, Oshlack A, Morley R, Carlin JB, Saffery R, Smyth GK, Craig JM: Expression discordance of monozygotic twins at birth: effect of intrauterine environment and a possible mechanism for fetal programming. *Epigenetics* 2011, **6**:579-592.
7. Gordon L, Joo EJ, Powell JE, Ollikainen M, Novakovic B, Li X, Andronikos R, Cruickshank MN, Conneely KN, Smith AK, Alisch RS, Morley R, Visscher PM, Craig JM, Saffery R: Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental/genetic factors subject to tissue-specific influence. *Genome Res* 2012, **22**:1395-1406.
8. Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J: A longitudinal study of epigenetic variation in twins. *Epigenetics* 2010, **5**:516-526.
9. Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, Kalidindi S, Picchioni M, Kravariti E, Touloupoulou T, Murray RM, Mill J: Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum Mol Genet* 2011, **20**:4786-4796.
10. Gervin K, Hammero M, Akselsen HE, Moe R, Nygard H, Brandt I, Gjessing HK, Harris JR, Undlien DE, Lyle R: Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome Res* 2011, **21**:1813-1821.
11. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, Kahn RS, Ophoff RA: The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One* 2009, **4**: e6767.
12. Rakan V, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD: Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 2010, **20**:434-439.

13. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, Dermitzakis ET, McCarthy MI, Mill J, Spector TD, Deloukas P: **Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population.** *PLoS Genet* 2012, **8**:e1002629.
14. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, Kocjan G, Gayther SA, Jones A, Beck S, Wagner W, Laird PW, Jacobs IJ, Widschwendter M: **Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer.** *Genome Res* 2010, **20**:440-446.
15. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, Vilain E: **Epigenetic predictor of age.** *PLoS One* 2011, **6**:e14821.
16. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT: **Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context.** *PLoS Genet* 2009, **5**:e1000602.
17. Hernandez MI, Mericq V: **Metabolic syndrome in children born small-for-gestational age.** *Arquivos Brasileiros Endocrinologia Metabologia* 2011, **55**:583-589.
18. Novakovic B, Gordon L, Wong NC, Moffett A, Manuelpillai U, Craig JM, Sharkey A, Saffery R: **Wide ranging DNA methylation differences of primary trophoblast cell populations and derived-cell lines: implications and opportunities for understanding trophoblast function.** *Mol Hum Reprod* 2011, **17**:344-353.
19. Schroeder JW, Conneely KN, Cubells JC, Kilaru V, Newport DJ, Knight BT, Stowe ZN, Brennan PA, Krushkal J, Tyllavsky FA, Taylor RN, Adkins RM, Smith AK: **Neonatal DNA methylation patterns associate with gestational age.** *Epigenetics* 2011, **6**:1498-1504.
20. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST: **Age-associated DNA methylation in pediatric populations.** *Genome Res* 2012, **22**:623-632.
21. Siegmund KD, Connor CM, Campan M, Long TI, Weisenberger DJ, Biniszkiwicz D, Jaenisch R, Laird PW, Akbarian S: **DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons.** *PLoS One* 2007, **2**:e895.
22. Yuen RK, Neumann SM, Fok AK, Penaherrera MS, McFadden DE, Robinson WP, Kobor MS: **Extensive epigenetic reprogramming in human somatic tissues between fetus and adult.** *Epigenet Chromatin* 2011, **4**:7.
23. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ, Puca AA, Sayols S, Pujana MA, Serra-Musach J, Iglesias-Platas I, Formiga F, Fernandez AF, Fraga MF, Heath SC, Valencia A, Gut IG, Wang J, Esteller M: **Distinct DNA methylomes of newborns and centenarians.** *Proc Natl Acad Sci USA* 2012, **109**:10522-10527.
24. Martino DJ, Tulic MK, Gordon L, Hodder M, Richman T, Metcalfe J, Prescott SL, Saffery R: **Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans.** *Epigenetics* 2011, **6**:1085-1094.
25. Martino D, Maksimovic J, Joo JH, Prescott SL, Saffery R: **Genome-scale profiling reveals a subset of genes regulated by DNA methylation that program somatic T-cell phenotypes in humans.** *Genes Immun* 2012, **13**:388-398.
26. Wang D, Liu X, Zhou Y, Xie H, Hong X, Tsai HJ, Wang G, Liu R, Wang X: **Individual variation and longitudinal pattern of genome-wide DNA methylation from birth to the first two years of life.** *Epigenetics* 2012, **7**:594-605.
27. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, Fallin MD: **Personalized epigenomic signatures that are stable over time and covary with body mass index.** *Sci Transl Med* 2010, **2**:49ra67.
28. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrom TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V, Feinberg AP: **Intra-individual change over time in DNA methylation with familial clustering.** *JAMA* 2008, **299**:2877-2883.
29. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A: **Decline in genomic DNA methylation through aging in a cohort of elderly subjects.** *Mech Ageing Dev* 2009, **130**:234-239.
30. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heinzel-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M: **Epigenetic differences arise during the lifetime of monozygotic twins.** *Proc Natl Acad Sci USA* 2005, **102**:10604-10609.
31. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE: **Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus.** *Hum Mol Genet* 2007, **16**:547-554.
32. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R: **High density DNA methylation array with single CpG site resolution.** *Genomics* 2011, **98**:288-295.
33. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M: **Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.** *Epigenetics* 2011, **6**:692-702.
34. Saffery R, Morley R, Carlin JB, Joo JH, Ollikainen M, Novakovic B, Andronikos R, Li X, Loke YJ, Carson N, Wallace EM, Umstad MP, Permezel M, Galati JC, Craig JM: **Cohort Profile: The Peri/post-natal Epigenetic Twins Study.** *Int J Epidemiol* 2012, **41**:55-61.
35. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F: **Evaluation of the Infinium Methylation 450K technology.** *Epigenomics* 2011, **3**:771-784.
36. Maksimovic J, Gordon L, Oshlack A: **SWAN: Subset quantile Within-Array Normalization for Illumina Infinium HumanMethylation450 BeadChips.** *Genome Biol* 2012, **13**:R44.
37. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
38. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL: **Genome-wide methylation profiling using Infinium assay.** *Epigenomics* 2009, **1**:177-200.
39. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA: **Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies.** *Int J Epidemiol* 2012, **41**:200-209.
40. Salsbery PJ, Reagan PB: **Effects of heritability, shared environment, and nonshared intrauterine conditions on child and adolescent BMI.** *Obesity (Silver Spring)* 2010, **18**:1775-1780.
41. Plomin R: **Commentary: Why are children in the same family so different? Non-shared environment three decades later.** *Int J Epidemiol* 2011, **40**:582-592.
42. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z: **GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.** *BMC Bioinformatics* 2009, **10**:48.
43. Schneider E, Pliushch G, El Hajj N, Galetzka D, Puhl A, Schorsch M, Frauenknecht K, Riepert T, Tresch A, Muller AM, Coerd W, Zechner U, Haaf T: **Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns.** *Nucleic Acids Res* 2010, **38**:3880-3890.
44. Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, McRae AF, Visscher PM, Montgomery GW, Martin NG, Petronis A: **DNA methylation profiles in monozygotic and dizygotic twins.** *Nat Genet* 2009, **41**:240-245.
45. Murphy SK, Huang Z, Hoyo C: **Differentially methylated regions of imprinted genes in prenatal, perinatal and postnatal human tissues.** *PLoS One* 2012, **7**:e40924.
46. Thiede C, Prange-Krex G, Freiberg-Richter J, Bornhauser M, Ehninger G: **Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants.** *Bone Marrow Transplant* 2000, **25**:575-577.
47. Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Winingner M, Colantuoni C, Weinberger DR, Kleinman JE, Lipska BK: **DNA methylation signatures in development and aging of the human prefrontal cortex.** *Am J Hum Genet* 2012, **90**:260-272.
48. Novakovic B, Yuen RK, Gordon L, Penaherrera MS, Sharkey A, Moffett A, Craig JM, Robinson WP, Saffery R: **Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors.** *BMC Genomics* 2011, **12**:529.
49. Thompson RF, Atzmon G, Gheorghie C, Liang HQ, Lowes C, Greally JM, Barzilai N: **Tissue-specific dysregulation of DNA methylation in aging.** *Ageing Cell* 2010, **9**:506-518.
50. Liang P, Song F, Ghosh S, Morien E, Qin M, Mahmood S, Fujiwara K, Igarashi J, Nagase H, Held WA: **Genome-wide survey reveals dynamic**

- widespread tissue-specific changes in DNA methylation during development. *BMC Genomics* 2011, **12**:231.
51. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature* 2008, **454**:766-770.
 52. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, Chen M, He Y, Jin X, Zhang Q, Yu C, Zhou G, Huang Y, Cao H, Zhou X, Guo S, Hu X, Li X, Kristiansen K, Bolund L, Xu J, Wang W, Yang H, et al: **The DNA methylome of human peripheral blood mononuclear cells.** *PLoS Biol* 2010, **8**:e1000533.
 53. Laurent L, Wong E, Li G, Huynh T, Tsigirig A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei CL: **Dynamic changes in the human methylome during differentiation.** *Genome Res* 2010, **20**:320-331.
 54. Gentilini D, Mari D, Castaldi D, Remondini D, Ogliaari G, Ostan R, Bucci L, Sirchia SM, Tabano S, Cavagnini F, Monti D, Franceschi C, Di Blasio AM, Vitale G: **Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring.** *Age (Dordr)* 2012.
 55. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA: **Aging effects on DNA methylation modules in human brain and blood tissue.** *Genome Biol* 2012, **13**:R97.
 56. Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, Moore M, Longo DL, Cookson MR, Traynor BJ, Singleton AB: **Distinct DNA methylation changes highly correlated with chronological age in the human brain.** *Hum Mol Genet* 2011, **20**:1164-1172.
 57. Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E, Oyolu CB, Schroth GP, Absher DM, Baker JC, Myers RM: **Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver.** *Genome Res* 2009, **19**:1044-1056.
 58. Kupfer DM, White VL, Jenkins MC, Burian D: **Examining smoking-induced differential gene expression changes in buccal mucosa.** *BMC Med Genomics* 2010, **3**:24.
 59. Sajjani MR, Patel AK, Bhatt VD, Tripathi AK, Ahir VB, Shankar V, Shah S, Shah TM, Koringa PG, Jakhesara SJ, Joshi CG: **Identification of novel transcripts deregulated in buccal cancer by RNA-seq.** *Gene* 2012, **507**:152-158.
 60. Ollikainen M, Smith KR, Joo EJ, Ng HK, Andronikos R, Novakovic B, Abdul Aziz NK, Carlin JB, Morley R, Saffery R, Craig JM: **DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome.** *Hum Mol Genet* 2010, **19**:4176-4188.
 61. Talens RP, Christensen K, Putter H, Willemssen G, Christiansen L, Kremer D, Suchiman HE, Slagboom PE, Boomsma DI, Heijmans BT: **Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs.** *Aging Cell* 2012, **11**:694-703.
 62. Martin GM: **Epigenetic drift in aging identical twins.** *Proc Natl Acad Sci USA* 2005, **102**:10413-10414.
 63. Petronis A: **Epigenetics and twins: three variations on the theme.** *Trends Genet* 2006, **22**:347-350.
 64. Whitelaw NC, Chong S, Whitelaw E: **Tuning in to noise: epigenetics and intangible variation.** *Dev Cell* 2010, **19**:649-650.
 65. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K: **Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates.** *Mol Cell* 2013, **49**:359-367.
 66. Pujadas E, Feinberg AP: **Regulated noise in the epigenetic landscape of development and disease.** *Cell* 2012, **148**:1123-1131.
 67. Touwslager RN, Gerver WJ, Mulder AL, Jansen AJ, de Bruin R: **Longitudinal growth during the first years of life: what is normal?** *Hormone Res* 2008, **70**:273-277.
 68. Barnett AG, van der Pols JC, Dobson AJ: **Regression to the mean: what it is and how to deal with it.** *Int J Epidemiol* 2005, **34**:215-220.
 69. Wilson RS: **Twin growth: initial deficit, recovery, and trends in concordance from birth to nine years.** *Ann Hum Biol* 1979, **6**:205-220.
 70. Keith L, Machin G: **Zygosity testing. Current status and evolving issues.** *J Reprod Med* 1997, **42**:699-707.
 71. Martin N, Boomsma D, Machin G: **A twin-pronged attack on complex traits.** *Nat Genet* 1997, **17**:387-392.
 72. Waddington CH: **Canalization of development and genetic assimilation of acquired characters.** *Nature* 1959, **183**:1654-1655.
 73. Waddington CH: *The Strategy of the Genes* London: Geo Allen and Unwin; 1957.
 74. R_Development_Core_Team: *R: A Language and Environment for Statistical Computing* Vienna, Austria: R Foundation for Statistical Computing; 2009.
 75. Hansen DK, Aryee MJ: **Analyze Illumina's 450k methylation arrays. R package version 1.6.0.** [http://bioconductor.org/packages/release/bioc/html/minfi.html].
 76. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**:R80.
 77. Du P, Kibbe WA, Lin SM: **lumi: a pipeline for processing Illumina microarray.** *Bioinformatics* 2008, **24**:1547-1548.
 78. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM: **Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.** *BMC Bioinformatics* 2011, **11**:587.
 79. Smyth GK: **Limma: linear models for microarray data.** In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Edited by: Gentleman R, Carey V, Dudoit S, Huber W, Irizarry R. New York: Springer; 2005:397-420.
 80. Docherty SJ, Davis OS, Haworth CM, Plomin R, Mill J: **DNA methylation profiling using bisulfite-based epityping of pooled genomic DNA.** *Methods* 2010, **52**:255-258.
 81. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc* 1995, **57**:289-300.
 82. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabuncyan S, Feinberg AP: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.** *Nat Genet* 2009, **41**:178-186.
 83. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nat Genet* 2009, **41**:1350-1353.
 84. Lian H, Thompson WA, Thurman R, Stamatoyannopoulos JA, Noble WS, Lawrence CE: **Automated mapping of large-scale chromatin structure in ENCODE.** *Bioinformatics* 2008, **24**:1911-1916.
 85. Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, Bodine DM, McKay RD, Chenoweth JG, Tesar PJ, Furey TS, Ren B, Weng Z, Crawford GE: **Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome.** *PLoS Genet* 2007, **3**:e136.
 86. Aryee MJ, Wu Z, Ladd-Acosta C, Herb B, Feinberg AP, Yegnasubramanian S, Irizarry RA: **Accurate genome-scale percentage DNA methylation estimates from microarray data.** *Biostatistics* 2011, **12**:197-210.
 87. Barton G, Abbott J, Chiba N, Huang DW, Huang Y, Krznaric M, Mack-Smith J, Saleem A, Sherman BT, Tiwari B, Tomlinson C, Aitman T, Darlington J, Game L, Sternberg MJ, Butcher SA: **EMAAS: an extensible grid-based rich internet application for microarray data analysis and management.** *BMC Bioinformatics* 2008, **9**:493.
 88. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions.** *Nat Biotechnol* 2010, **28**:495-501.

doi:10.1186/gb-2013-14-5-r42

Cite this article as: Martino et al.: Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biology* 2013 **14**:R42.