

# Guthrie card methylomics identifies temporally stable epialleles that are present at birth in humans

Huriya Beyan,<sup>1,6</sup> Thomas A. Down,<sup>2,6</sup> Sreeram V. Ramagopalan,<sup>1,3,6</sup> Kristina Uvebrant,<sup>4</sup> Anita Nilsson,<sup>4</sup> Michelle L. Holland,<sup>1</sup> Carolina Gemma,<sup>1</sup> Gavin Giovannoni,<sup>1</sup> Bernhard O. Boehm,<sup>5</sup> George C. Ebers,<sup>3</sup> Åke Lernmark,<sup>4</sup> Corrado M. Cilio,<sup>4</sup> R. David Leslie,<sup>1,7</sup> and Vardhman K. Rakan<sup>1,7</sup>

<sup>1</sup>The Blizzard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, United Kingdom; <sup>2</sup>The Gurdon Institute and Department of Genetics, University of Cambridge, Cambridge CB2 1QN, United Kingdom; <sup>3</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, Oxford OX3 7BN, United Kingdom; <sup>4</sup>Department of Clinical Sciences, Lund University, Skåne University Hospital SUS, 205 02 Malmö, Sweden; <sup>5</sup>Division of Endocrinology and Diabetes, Department of Internal Medicine I, University Medical Center Ulm and Center of Excellence "Metabolic Disorders" Baden-Württemberg, 89070 Ulm, Germany

A major concern in common disease epigenomics is distinguishing causal from consequential epigenetic variation. One means of addressing this issue is to identify the temporal origins of epigenetic variants via longitudinal analyses. However, prospective birth-cohort studies are expensive and time consuming. Here, we report DNA methylomics of archived Guthrie cards for the retrospective longitudinal analyses of in-utero-derived DNA methylation variation. We first validate two methodologies for generating comprehensive DNA methylomes from Guthrie cards. Then, using an integrated epigenomic/genomic analysis of Guthrie cards and follow-up samplings, we identify interindividual DNA methylation variation that is present both at birth and 3 yr later. These findings suggest that disease-relevant epigenetic variation could be detected at birth, i.e., before overt clinical disease. Guthrie card methylomics offers a potentially powerful and cost-effective strategy for studying the dynamics of interindividual epigenomic variation in a range of common human diseases.

[Supplemental material is available for this article.]

There is increasing interest in exploring the possibility that interindividual epigenetic variation plays a role in common human disease etiology (for review, see Petronis 2010; Rakan et al. 2011). Consequently, epigenome-wide association studies (EWASs), analogous to genome-wide association studies (GWASs) for disease-associated genetic variants (Rakan et al. 2011), are currently being conducted to systematically uncover epigenetic variants for a range of different diseases ([www.roadmapepigenomics.org/participants](http://www.roadmapepigenomics.org/participants) lists some of the current NIH-funded EWASs). However, the disease process itself can also induce epigenetic changes. Therefore, an EWAS design that just compares post-disease affected with unaffected individuals will not distinguish causal from consequential disease-associated epigenetic variation. The inability to make this distinction will ultimately prove to be a significant barrier to establishing the etiological role or medical value for any disease-associated epigenetic variants.

One way of addressing this issue is to determine whether the epigenetic variants are detectable before overt disease. Although such variants could arise at any point during an individual's lifespan (Fraga et al. 2005), in-utero development is considered to be a period during which the epigenome is particularly susceptible to stochastic and/or environmental factors such as infection, poor diet, or stress (for review, see Warner and Ozanne 2010). A variety

of common diseases have been proposed to have a component of in-utero origins including asthma, metabolic conditions including type 2 diabetes, and autoimmune disorders like multiple sclerosis (Heijmans et al. 2008; Warner and Ozanne 2010). Hence, being able to determine epigenomic profiles at birth, and by implication before overt disease, would be invaluable for elucidating the origins and potentially the etiological role of disease-associated epigenetic variants. In this regard, prospective birth-cohort studies are very useful but, unfortunately, are expensive, time-consuming, and difficult to establish.

In 1963, Robert Guthrie proposed spotting a small amount of neonatal blood onto filter paper for the large-scale biochemical screening of various diseases such as phenylketonuria (Guthrie and Susi 1963). "Guthrie cards" are now routinely created by major national health authorities; e.g., since 2000, over 90% of all neonates in the United States had Guthrie cards made shortly after birth (Pollitt 2009). Remarkably, in many cases Guthrie cards are being stored indefinitely. This led us to consider genome-wide DNA methylation profiling—DNA methylomics—of Guthrie cards as a means of investigating the origins of disease-associated epigenetic variation. Crucially, this longitudinal investigation could be done retrospectively at a fraction of the cost and time investment required for prospective studies.

## Results

Our first objective was to confirm that Guthrie cards could be used to generate high-quality DNA methylomes. We obtained Guthrie cards from the Diabetes Prediction in Skåne (DiPiS) study (Larsson et al. 2004). These cards were spotted with cord blood in September

<sup>6</sup>These authors contributed equally to this work.

<sup>7</sup>Corresponding authors

E-mail [v.rakan@qmul.ac.uk](mailto:v.rakan@qmul.ac.uk)

E-mail [r.d.g.leslie@qmul.ac.uk](mailto:r.d.g.leslie@qmul.ac.uk)

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.134304.111>.

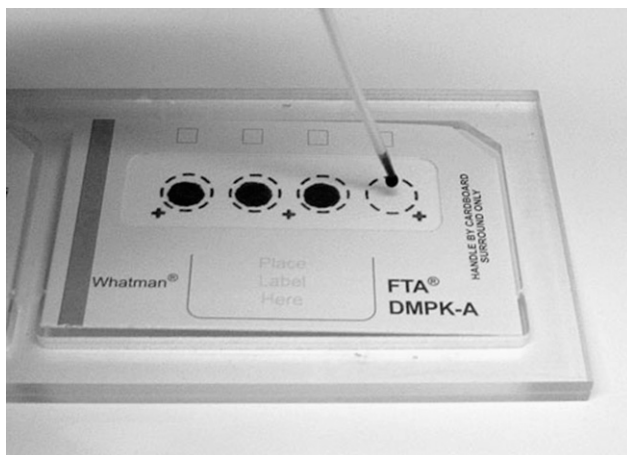
2000 (Methods). Of several different protocols, we found the Gensolve kit provided the highest genomic DNA (gDNA) yields from Guthrie cards (data not shown). A typical Guthrie card consists of four ~6-mm diameter blood spots, and we routinely obtained ~200 ng of gDNA per spot (Fig. 1). Although Guthrie card gDNA was of a slightly lower molecular weight compared with gDNA from freshly sampled cord blood, significant degradation was minimal (Supplemental Fig. 1). We then validated two different platforms for DNA methylomics, each offering different advantages: Illumina 450K arrays (Bibikova et al. 2011; Sandoval et al. 2011) and Methylated DNA Immunoprecipitation combined with HTS (MeDIP-seq) (Weber et al. 2005; Down et al. 2008).

The Illumina450K array is a genome-scale platform that provides bisulfite conversion-based, single-base resolution methylation measurements at ~450,000 different cytosines associated with a range of genomic features such as promoters, enhancers, and CpG islands (CGIs) (Sandoval et al. 2011). We generated biological duplicate Illumina450K profiles for Guthrie cards, freshly sampled cord blood, and mature sperm, all from unrelated individuals (Supplemental Table 1). Comparison of Illumina450K-based “fresh blood vs. sperm” tissue-specific differentially methylated regions (tDMRs) with “blood vs. sperm” tDMRs identified by bisulfite PCR sequencing in the Human Epigenome Project (Eckhardt et al. 2006) revealed a high correlation ( $R^2 = 0.90$ , Pearson’s) (Fig. 2A). Importantly, a comparison between Illumina450K-based fresh blood and Guthrie card profiles also revealed an excellent correlation ( $R^2 = 0.99$ , Pearson’s) (Fig. 2B). However, in both cases, the correlation was significantly weaker for tDMR methylation differences of <20% ( $R^2$  for <20% methylation differences = 0.50, Pearson’s), and thus we set a conservative minimum threshold of  $\geq 20\%$  methylation difference for calling intersample DMRs in the analyses described below (in the rest of the manuscript, when we state “% methylation difference,” we mean an absolute difference in the beta value  $\times 100$ ). To further confirm the quality of Guthrie card methylomes, we compared canonical gene-associated DNA methylation profiles for both CGI and non-CGI associated genes between Guthrie cards and fresh blood and found the profiles to be virtually identical for the two different types of samples (Supplemental Fig. 2).

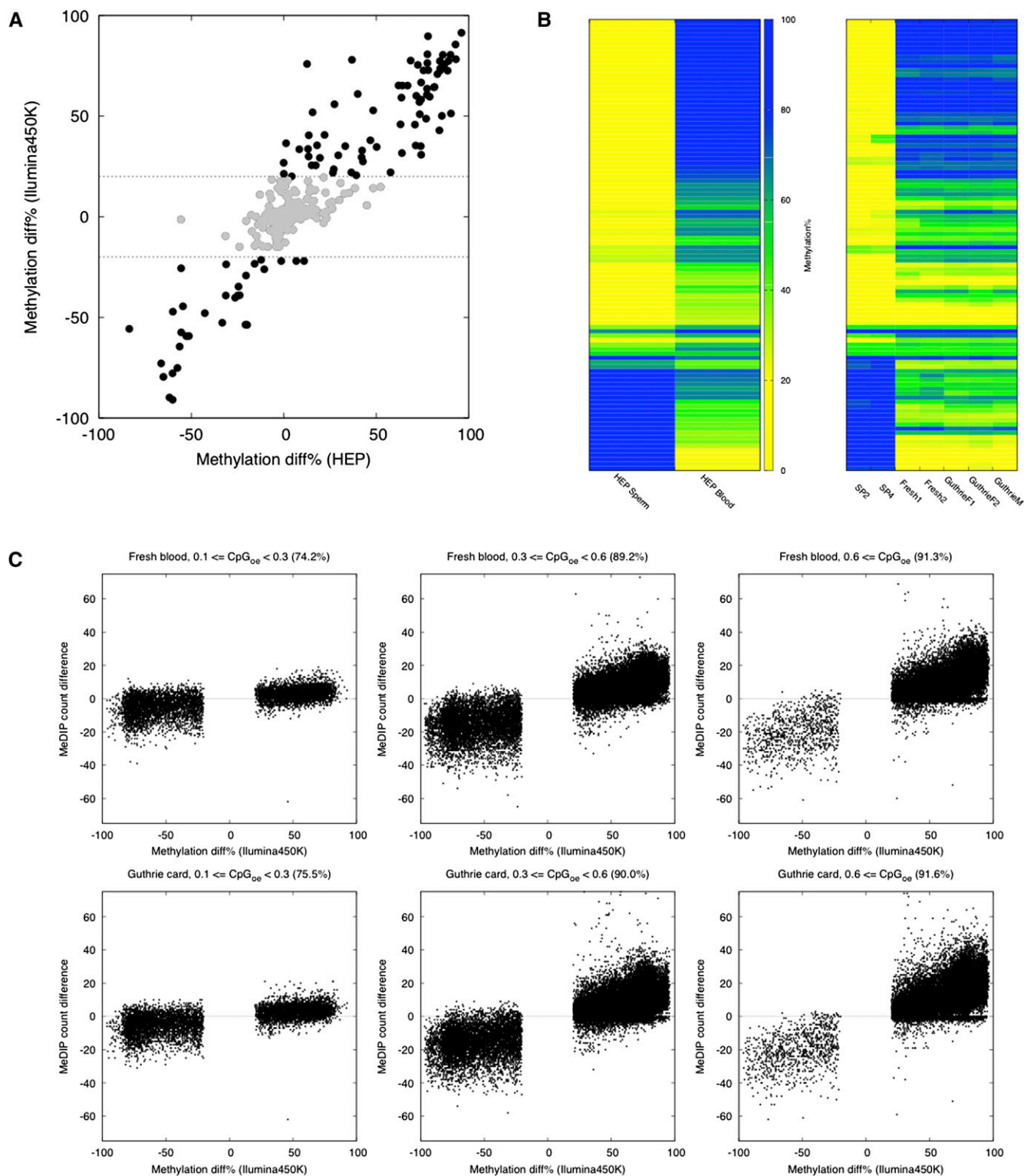
In MeDIP-seq, immunoprecipitated methylated DNA fragments are subjected to HTS, representing a cost-effective strategy for generating whole-genome DNA methylomes at ~50 bp resolution (Down et al. 2008). For Guthrie cards analysis, we developed

a MeDIP-seq protocol for 200 ng of gDNA (Methods). We also generated standard MeDIP-seq libraries using 2  $\mu\text{g}$  of DNA from fresh cord blood and sperm (Supplemental Table 1). The MeDIP-seq data were then compared with the Illumina450K-based profiles described above (Fig. 2C). Because MeDIP enrichment of a genomic region is dependent on the density of methylated CpG sites, standard measures of correlation are not suitable, and instead we calculated an “agreement” score: the percentage of Illumina450K-based DMRs that show the same directional differences in MeDIP-seq read counts between the two samples under consideration. Very good agreement scores (75%–93%) were observed between Illumina450K and MeDIP-seq data sets across a range of CpG densities. A similar analysis using Guthrie card vs. sperm MeDIP-seq-based tDMRs yielded equally strong agreement with Illumina450K data (Fig. 2C).

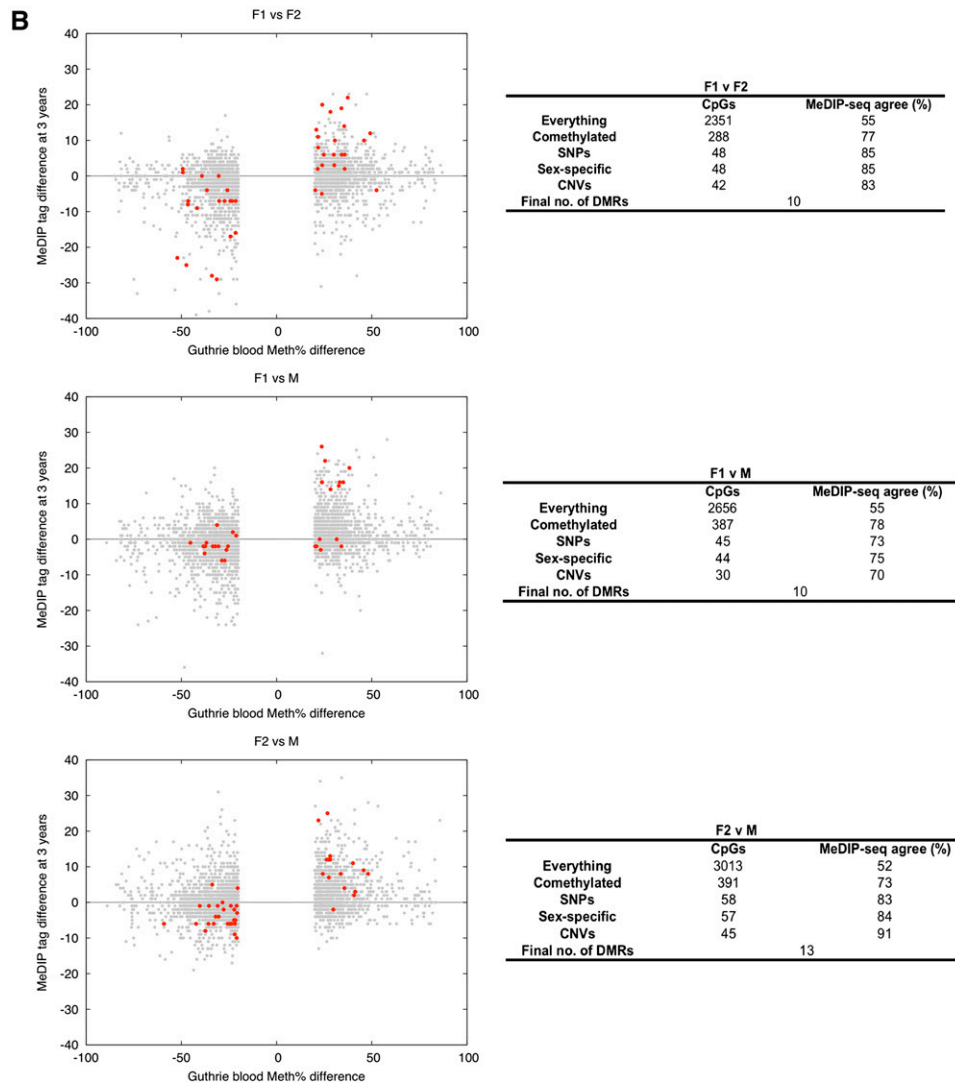
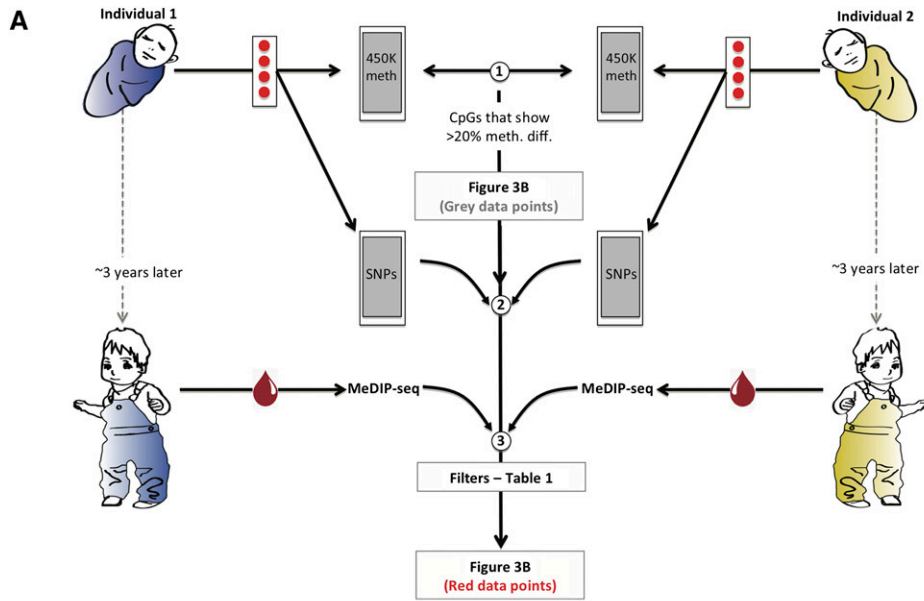
Our second objective was to determine the potential utility of Guthrie cards for studying human interindividual epigenetic variation. If in-utero-derived epigenomic perturbations can impact on phenotypic outcomes in later life, then there must be some interindividual epigenetic variants that are detectable at birth and temporally stable. To obtain evidence for the existence of such epigenetic variants, we conducted an integrated methylomic/genomic analysis of Guthrie cards and follow-up samplings as outlined in Figure 3A. We obtained Guthrie cards spotted with cord-blood from three unrelated individuals (two females and one male) from the DiPiS study and generated Illumina450K-based DNA methylomes. For each of the three different pairwise comparisons, we found ~2000–3000 CpG sites that displayed interindividual DNA methylation differences of >20% (Fig. 3B) (sex chromosomes were excluded from all analyses). We then filtered these lists to reduce the influence of artifactual methylation measurements and interindividual genetic variation (Table 1). To this end, we first generated genome-wide single nucleotide polymorphism (SNP) profiles for all three individuals using their Guthrie card DNA. SNP profiling was done using the Illumina Omni2.5S array that contains ~2.5 million SNPs with a minor allele frequency of down to 1%. Using these data, we applied three main filters to each of the three different pairwise lists of interindividual single CpG methylation differences: (1) excluded those CpGs that overlapped single nucleotide polymorphisms (SNPs) or were located within a copy number variant (CNV). This eliminates artifactual measurements that are not methylation differences, but rather due to the creation/destruction of CpGs or differences in copy number, an issue rarely addressed in complex disease epigenomics. (2) As the biological relevance of methylation differences limited to single CpG sites is presently unclear, for the second filter we specified that the CpG must exist as part of a differentially methylated region (DMR), i.e., neighboring CpGs within a 250-bp window of the index CpG site must also display >20% methylation differences in the same direction. (3) Recently, Bell and colleagues identified 180 different CpGs in the human genome whose methylation state in lymphoblastoid cells is influenced by genetic variation (Bell et al. 2011). Crucially, in the vast majority of cases they found the genetic influence to act in *cis* and <5 kb from the CpG site of interest. Therefore, for the third filter, we retained only those DMRs that are located within a  $\geq 10$ -kb window that does not contain any SNP differences, as measured by the Illumina Omni2.5S SNP array in the pairwise comparison under consideration. This reduces the likelihood of interindividual genetic differences being the sole drivers of the DMRs (although in a few cases rare genetic variants missing from the SNP arrays would be overlooked). Overall, application of the filters resulted in a final



**Figure 1.** An example of the type of Guthrie cards used in our study (image from Ake Lernmark).



**Figure 2.** Array and sequencing-based DNA methylomics of Guthrie cards. (A) Comparison of Illumina450K-based “fresh cord blood vs. sperm” tDMRs with “blood vs. sperm” tDMRs identified in the Human Epigenome Project (HEP) (Eckhardt et al. 2006). Plotted are 468 genomic regions (~200–300 bp in length) common between the HEP and the Illumina450K data sets. Illumina450K-based “fresh cord blood vs. sperm” tDMRs that display <20% methylation differences are shown in gray.  $R^2 = 0.90$  (overall),  $R^2 = 0.50$  (<20% methylation), both Pearson’s. (B) A comparison of fresh cord blood and Guthrie card Illumina450K profiles with HEP data. Shown is methylation data for regions called “blood vs. sperm” tDMRs in the HEP. Each column represents profiles from a separate individual, and 106 different tDMRs are shown. (SP) Sperm. (C) Normalized MeDIP-seq read counts for the “blood vs. sperm” tDMRs identified by the use of Illumina450K-based fresh blood (top) or Guthrie cards (bottom) across a range of CpG densities. Only >20% tDMR methylation differences are shown. The percentage agreement is indicated in each panel. Number of different CpGs represented in each panel: Low CpG,  $n = 6404$ ; mid-CpG,  $n = 28,384$ ; high CpG,  $n = 18,135$ . Note the numbers are the same for the “Fresh blood” and “Guthrie card” data sets as the differentially methylated CpG sites were called on the Illumina450K arrays.



**Figure 3.** Identification at birth of temporally stable individual-specific DNA methylation variants. (A) Experimental strategy schematic: (1) Guthrie card Illumina450K profiles were generated and >20% interindividual pairwise methylation differences called (sex chromosome data were excluded from all analyses); (2) SNP data were also generated from Guthrie card DNA; (3) a variety of filters (detailed in Table 1) were applied to the Illumina450K profiles, and then validation was performed using MeDIP-seq profiles of blood sampled when the individuals were 3 yr old. (B) Identification of temporally stable interindividual DMRs that exist at birth. (Gray points) Single CpG interindividual differences of >20%; (red points) the filtered set. The number of CpGs remaining after each filter is indicated in the adjoining tables. The final sets of CpGs (red points in the figure) were grouped into “DMRs”—differentially methylated regions. Binomial  $P$ -values (calculated at a per-DMR level) for the “Agreement scores” between Illumina450K and MeDIP-seq data are: F1 vs. F2:  $P = 0.01$ , F1 vs. M:  $P = 0.11$ , F2 vs. M:  $P = 0.0003$ .

**Table 1.** Filters applied to the DMRs

chrX and Y probes	To exclude sex-specific DNA methylation differences
CpG SNPs	These would create/destroy CpG sites and hence would lead to artifactual interindividual methylation differences.
CNVs	These would also lead to artifactual methylation differences measured by the arrays (e.g., increased copy number would show up as increased methylation).
Comethylation	All CpGs within a 250-bp window must display >20% methylation differences in the same direction. The window must have at least 2 CpGs. This will eliminate single CpG differences that are either likely to be artifactual and/or have minimal functional impact.
Genetic identity	No interindividual SNP differences for at least 5 kb upstream of and downstream from the DMR. Furthermore, there must be data from at least 5 SNPs on the array in this region.

set of 10–13 different interindividual DMRs per pairwise comparison (Fig. 3B; Table 2). Given the stringency of our filters, this is almost certainly an underestimate of the level of interindividual DNA methylation variation that exists between any two unrelated human individuals.

Then, to establish the temporal stability of the DMRs found using Guthrie cards (i.e., at birth), we generated MeDIP-seq profiles of whole blood sampled from the same individuals when they were 3 yr old (all individuals were disease-free at the time of sampling). Since we did not have access to 500 ng of gDNA required to perform 450K analysis on the follow-up samples, we used the MeDIP-seq method for 200 ng of gDNA as described above. Given the differences in platform-specific biases, the eventual results would, if anything, be an underestimate of the number of temporally stable epialleles that exist in the individuals under study. MeDIP-seq sequencing statistics are provided in the Supplemental section. Comparison between the final set of “at birth” Illumina450K-based interindividual DMRs with MeDIP-seq profiles of the same individuals at 3 yr old yielded agreement scores ranging from 70% to 91% (Fig. 3B). This demonstrates, for the first time, the existence of temporally stable interindividual epigenetic variation that is present at birth. These DMRs were found across the genome associated with a variety of genomic elements, including promoters and CGIs, but also intergenic regions (Table 2). Interestingly, a few DMRs were located within genes associated with human diseases, e.g., FZD1 expression is down-regulated in ulcerative colitis (Hughes et al. 2011) and NFIA has been associated with celiac disease (Dubois et al. 2010). It is also worth noting the large difference in agreement scores between the unfiltered (52%–55% i.e., marginally better than random) and post-filtered data (70%–91%). This suggests that future epigenomic investigations should integrate epigenetic and genetic variation data to optimize the biological relevance and reduce false-positive “epigenetic” discoveries.

## Discussion

A significant challenge in EWASs will be distinguishing causal from consequential epigenetic variation, and using solely “case vs. control” study designs will not help to address this issue. Rather, as has been recently discussed (Rakyan et al. 2011), effective EWASs will need to combine different types of cohorts, among which longitudinal cohorts in particular will be invaluable. Guthrie card DNA methylomics therefore has the potential to become a valuable component of EWASs for a range of common diseases.

Guthrie cards could offer several advantages over large-scale prospective studies. First, in most cases they are created and

stored by national health authorities, e.g., [www.newbornbloodspot.screening.nhs.uk](http://www.newbornbloodspot.screening.nhs.uk). Second, if Guthrie cards are combined with an initial post-disease profiling of cases vs. controls, then only affected individuals and a suitable number of controls need to be included in the retrospective longitudinal analyses, whereas in prospective cohorts it is very difficult to predict which individuals will develop the disease of interest, necessitating very large sample numbers. Third, although at present Guthrie cards are available for individuals <20 yr old, this already represents millions of individuals, many of whom have developed a disease (e.g., both

type 1 diabetes and asthma represent childhood-onset diseases), and these numbers will only increase. In terms of potential limitations, Guthrie cards are restricted to whole blood that represents a heterogeneous population of cells. But this is also typically the case in large-scale prospective studies in which, for logistical reasons, it is only possible to sample tissues such as whole blood or buccals. However, since in-utero events are thought to impact on more than one tissue, Guthrie cards are likely to be informative for the epigenetics of a range of diseases that affect multiple blood subtypes and possibly even non-blood tissues. Although one could argue that prospective studies are likely to obtain more detailed information about life-style factors, Guthrie cards could still be used to establish whether disease-associated DNA methylation variants are present at birth, providing a crucial starting point for further research, including investigating the possibility of using epigenetic variants, in combination with genetics, as predictive markers. Finally, the vast numbers of Guthrie cards that currently exist provide a pool of material to study a large range of diseases. It is also worth noting that spotting blood on filter paper is not limited to neonatal screening, but could also be more widely applied in home/field-based settings.

Using Guthrie cards, we were able to identify temporally stable epialleles that are present at birth in humans, demonstrating a key application of Guthrie card methylomics. The identified DMRs described above must have been present in-utero since they were identified at birth. Given the filters we applied to the data, the DMRs are more likely due to environmental or stochastic interindividual differences within the in utero milieu (e.g., differences in maternal diet) rather than solely genetic differences. Indeed, even though ~95% of the CpGs assayed by Illumina27K arrays are also represented on the Illumina450K array (Bell et al. 2011), none of the CpGs they identified, at which the methylation state is influenced by *cis*-genetic variation, overlap our final filtered list of DMRs. However, to unequivocally rule out genetic differences, extensive resequencing of the DMRs would be required, which was not feasible for our proof-of-principle study, but should be performed in future larger-scale EWASs. Although in this study we focused on identifying epigenetic alterations that are less likely to be due to genetic heterogeneity, Guthrie card methylomics will also be useful for studying genotype–epigenotype interactions. Recently, it has been shown that genetic influences on epigenetic states in the human genome can be probabilistic, i.e., there isn't a strict one-to-one relationship between genotype and epigenotype, but rather a certain probability at which a given epigenetic state will arise (Feinberg and Irizarry 2010). Such regions have been termed “variably methylated regions” or VMRs (Feinberg and Irizarry 2010). VMRs

**Table 2.** Genomic characteristics of the interindividual DMRs

F1 vs. F2													
Chr	Start	End	CpG <sub>o/e</sub>	CpGs	Distance to nearest gene	Gene name	Location	Illumina450K		MeDIP-seq		Genetic identity	
								F1	F2	F1	F2	No. of SNPs	Length (bp)
2	130,617,902	130,618,144	0.59	2	0	CCDC74B	Promoter	41	13	5	1	15	132,212
5	1,647,282	1,647,676	0.84	4	38,174	Q7Z6L2_HUMAN	Intergenic	6	34	0	2	66	98,161
6	236,687	236,909	0.58	4	144	DUSP22	Promoter	54	11	9	5	9	21,713
6	32,597,779	32,597,941	0.77	2	0	HLA-DRB5	Exonic	55	77	9	5	9	50,611
6	32,633,783	32,634,238	0.49	7	20,286	HLA-DRB1	Intergenic	53	82	9	19	7	27,586
7	90,734,508	90,734,637	0.80	2	0	FZD1	Exonic	53	30	31	13	5	16,568
10	3,588,486	3,588,627	0.96	2	219,561	KLF6	Intergenic	57	78	7	19	53	62,081
12	31,163,381	31,163,386	0.55	2	0	NP_001073971.1	Exonic	37	13	9	2	11	43,211
16	87,935,749	87,935,904	0.51	4	0	ANKRD11	Intronic	78	56	17	8	22	30,772
17	76,189,287	76,189,311	0.49	2	0	RPTOR_HUMAN	Intronic	96	47	11	13	7	17,748

F1 vs. M													
Chr	Start	End	CpG <sub>o/e</sub>	CpGs	Distance to nearest gene	Gene name	Location	Illumina450K		MeDIP-seq		Genetic identity	
								F1	M	F1	M	No. of SNPs	Length (bp)
1	16,898,845	16,898,993	0.85	2	0	ESPNNP	Exonic	11	41	0	15	7	46,758
2	132,002,900	132,003,135	0.62	2	0	CCDC74A	Promoter	42	18	2	1	51	88,079
6	30,003,095	30,003,239	0.59	6	0	HLA-G	Intronic	64	38	2	0	8	15,991
6	32,659,927	32,660,130	0.78	10	0	HLA-DRB1	Exonic	62	78	25	22	17	24,352
10	3,588,486	3,588,627	0.96	2	219,561	KLF6	Intergenic	57	81	7	31	14	12,280
10	123,234,526	123,234,581	0.42	2	0	FGFR2	Intronic	74	42	8	5	47	36,679
14	105,163,026	105,163,196	0.37	2	23,240	LOC649910	Exonic	78	51	4	8	24	226,280
17	68,873,489	68,873,496	0.38	2	0	SDK2	Intronic	71	43	18	12	19	18,353
19	42,452,423	42,452,434	0.65	2	26,017	ZNF383	Intergenic	86	54	2	0	5	49,176

F2 vs. M													
Chr	Start	End	CpG <sub>o/e</sub>	CpGs	Distance to nearest gene	Gene name	Location	Illumina450K		MeDIP-seq		Genetic identity	
								F2	M	F2	M	No. of SNPs	Length (bp)
1	16,898,845	16,898,993	0.85	2	0	ESPNNP	Exonic	13	41	2	15	40	244,262
1	61,290,395	61,290,464	0.68	2	0	NFIA	Intronic	21	47	8	15	29	35,909
1	91,784,996	91,785,203	0.60	3	21,087	CDC7	Intergenic	80	54	29	20	13	14,588
3	198,190,026	198,190,139	0.69	2	9,925	PIGZ	Intergenic	44	13	5	1	16	17,087
4	75,066,510	75,066,693	0.59	7	0	PF4	Promoter	58	36	11	5	80	113,719
6	237,329	237,596	0.92	4	0	DUSP22	Promoter	8	43	2	15	14	36,281
6	32,598,399	32,598,422	0.32	2	0	HLA-DRB5	Intronic	42	83	1	9	9	50,611
6	139,054,553	139,054,685	0.50	2	615	Null	Promoter	78	43	6	5	8	11,476
7	90,734,508	90,734,637	0.80	2	0	FZD1	Exonic	30	55	13	37	75	108,395
11	36,379,031	36,379,191	0.46	3	0	NP_079117.2	Promoter	63	41	3	1	22	16,131
12	6,528,425	6,528,639	0.47	2	0	NP_542768.1	Intronic	64	41	13	8	52	46,668
14	105,163,026	105,163,196	0.37	2	23,240	LOC649910	Exonic	79	51	3	8	19	111,196
19	42,452,423	42,452,434	0.65	2	26,017	ZNF383	Intergenic	89	54	6	0	5	49,176

The "Illumina450K" and "MeDIP-seq" columns show %methylation and normalized read counts, respectively. In the "Genetic identity" columns, the extent of the genetic identity (in kilobase and number of SNPs) for the region containing the DMR is reported.

show similar properties to "metastable epialleles" previously described in mice (Morgan et al. 1999; Rakyant et al. 2003) and humans (Waterland et al. 2010), at which the establishment of the epigenetic state occurs during early development in a probabilistic manner. Therefore, if genetically influenced probabilistic establishment of epigenetic state occurs during embryogenesis, then knowing the epigenetic state at birth would be invaluable for understanding the biological impact and predicting the phenotypic outcomes of such genetic variants.

In summary, Guthrie cards will allow retrospective investigations into the temporal origins of epigenetic variants and potentially help elucidate their etiological role. Robert Guthrie

envisaged that such cards would eventually find many uses for understanding and ultimately preventing human diseases; the application to complex disease epigenomics would be a novel and powerful use of this remarkable resource.

## Methods

### Samples

Guthrie card and fresh cord blood samples were obtained from the DiPIS (Diabetes Prediction in Skåne) study (Supplemental Table 1). The cord blood-spot cards used in this study were created

in September 2000 and stored in the dark at ambient temperature (20–21°C) in individual page protectors. Umbilical cord blood was collected by venopuncture into BD Vacutainer EDTA tubes (BD), spotted onto filters grade 2992 (Schleicher & Schuell), and air-dried before storage. Fresh cord blood samples (from September, 2010) were collected into EDTA tubes at Malmö University Hospital, Lund University, and processed within 24 h. For the samplings of the 3-yr-old children, venous EDTA blood sample was obtained at a Community Health Center. The EDTA blood was immediately added to the Guthrie card in three droplets and allowed to dry before being mailed overnight in a page protector to the laboratory in Malmö, Sweden. The Guthrie cards were stored in the dark at room temperature before analysis within a month. The DNA from fresh cord blood was prepared using QIAamp DNA Blood Midi kit (Qiagen) according to the manufacturer's protocol. The Lund University Research Ethics Committee approved the study and written informed consent was obtained from participating mothers (DiPiS ethics approval no. Dnr 2009/244).

### Isolation of genomic DNA from Guthrie cards

Guthrie card gDNA was extracted using the GenSolve kit as per the manufacturer's instructions ([www.genvault.com/html/products](http://www.genvault.com/html/products)). Following extraction, the samples were further purified using the Qiagen QiaAmp mini columns and eluted in 100 µL of elution buffer. DNA concentration was determined using a Qubit instrument (Invitrogen).

### Generation and analysis of array- and HTS-based data

Illumina450K arrays were processed according to the manufacturer's instructions using 500 ng of DNA. The data were quantile normalized and resulting beta values were used in downstream analyses. MeDIP-seq libraries were prepared using a modified version of the protocol described in Reference 13 and sequenced on an Illumina GAIIX instrument (v4 chemistry). One lane of 36-bp paired-end sequencing was performed for each MeDIP-seq library. A detailed description is provided in the Supplemental Methods. MeDIP-seq reads were aligned and processed using bwa 0.5.9 (Li et al. 2009), samtools 0.1.7 (Li and Durbin 2009), and custom scripts (available upon request). After alignment and deduplication, we selected all read pairs with a mapping score  $\geq 10$ . Within each group of samples (Guthrie cards and fresh blood) we calculated histograms of read-pair spacings (insert sizes), then randomly discarded some reads from each library in order to perfectly match the insert-size distributions within the group. We then counted numbers of read pairs overlapping 500-bp windows centered around the target CpG site of each probe on the Illumina450K array. It is worth noting that we did not perform DMR calling per se on the MeDIP-seq data, but rather asked whether the difference between read counts in the two samples was positive or negative. So, read/tag depth only affects agreement score to the extent that signal/noise ratio will tend to be higher in a region with high read depths. Since MeDIP enrichment efficiency depends on the number of meCpGs per fragment, this will tend to be higher when methylation occurs in the context of CpG dense regions. Illumina Omni2.5S arrays ([www.illumina.com](http://www.illumina.com)) were processed according to the manufacturer's instructions using 200 ng of Guthrie card DNA.

### Statistical analyses

For all correlation analyses among the different methylomic data sets, Pearson's correlation coefficient was selected as it is a familiar descriptive measure of the similarity between two quantitative

data sets and has been previously used in the context of DNA methylomics (Down et al. 2008).

### Processing of HEP data

HEP bisulfite data for sperm, CD4+, and CD8+ T-cells were downloaded from the website. This includes amplicon coordinates (NCBI34) and methylation ratios for each CpG covered by each bisulfite PCR sequencing read of the HEP data set. Since some amplicons have low coverage, we selected only those amplicons which, for each of CD4+, CD8+, and Sperm, have at least two reads of data, and have coverage for at least five separate CpG sites. Also, to avoid double counting (i.e., where two HEP amplicons overlapped in our data set) we excluded the smaller of the overlapping amplicons from further analyses. For each amplicon that fulfilled these criteria, we took the mean of all CpG methylation ratios for each cell type. As an additional quality control measure (and also to exclude some HEP amplicons that are highly variable between different blood cell types) we discarded amplicons with a >10% methylation difference between CD4+ and CD8+ cells. We then took the mean of the CD4+ and CD8+ methylation ratios to give a combined "Blood" methylation value.

### Additional methods

Detailed methodology is described in the Supplemental Methods.

### Data access

The raw array and sequencing data from this study have been deposited in the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under accession no. GSE38128.

### Acknowledgments

We thank the midwives at Malmö University Hospital, Sweden, for their help in collecting cord blood and especially the participating mothers. This work was supported by the Knut and Alice Wallenberg Foundation. We acknowledge coinvestigators in the DiPiS study: C. Andersson, B. Lindberg, B. Lernmark, A. Carlsson, E. Cederwall, K. Larsson, J. Neiderud, B. Jönsson, S.A. Ivarsson, and H.E. Larsson. The DiPiS study was supported by the Swedish Research Council (K2011-54X-15312-07-6) and the Skåne County Council for Research and Development. V.K.R., C.G., and R.D.L. are supported by the BBSRC, UK (BB/H012494/1). R.D.L. is supported by Juvenile Diabetes Research Foundation International (JDRFI Award 5-2011-145). S.V.R. and G.C.E. are funded by the Multiple Sclerosis Society of the United Kingdom. S.V.R. and G.G. are funded by the Medical Research Council, United Kingdom (G0801976). V.K.R., R.D.L., A.L., and B.O.B. are supported by EU-FP7 "BLUEPRINT." R.D.L. was also supported by grants from the JDRF (nPOD: 25-2010-703) and Diabetes UK (10/0004107).

### References

- Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* **12**: R10. doi: 10.1186/gb-2011-12-1-r10.
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. 2011. High density DNA methylation array with single CpG site resolution. *Genomics* **98**: 288–295.
- Down TA, Rakyian VK, Turner DJ, Flicek P, Li H, Kulesha E, Gräf S, Johnson N, Herrero J, Tomazou EM, et al. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* **26**: 779–785.

- Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhemakova A, Heap GA, Adány R, Aromaa A, et al. 2010. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* **42**: 295–302.
- Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, et al. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* **38**: 1378–1385.
- Feinberg AP, Irizarry RA. 2010. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci* **1**: 1757–1764.
- Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, et al. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci* **102**: 10604–10609.
- Guthrie R, Susi A. 1963. A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics* **32**: 338–343.
- Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Siagboom PE, Lumey LH. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci* **105**: 17046–17049.
- Hughes KR, Sablitzky F, Mahida YR. 2011. Expression profiling of Wnt family of genes in normal and inflammatory bowel disease primary human intestinal myofibroblasts and normal human colonic crypt epithelial cells. *Inflamm Bowel Dis* **17**: 213–220.
- Larsson K, Elding-Larsson H, Cederwall E, Kockum K, Neiderud J, Siöbiad S, Lindberg B, Lernmark B, Cilio C, Ivansson SA, et al. 2004. Genetic and perinatal factors as risk for childhood type 1 diabetes. *Diabetes Metab Res Rev* **20**: 429–437.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Horner N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**: 314–318.
- Petronis A. 2010. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* **465**: 721–727.
- Pollitt RJ. 2009. Newborn blood spot screening: New opportunities, old problems. *J Inherit Metab Dis* **32**: 395–399.
- Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, Whitelaw E. 2003. Transgenerational inheritance of epigenetic states at the murine *Axin<sup>fl</sup>* allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci* **100**: 2538–2543.
- Rakyan VK, Down TA, Balding DJ, Beck S. 2011. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* **12**: 529–541.
- Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**: 692–702.
- Warner MJ, Ozanne SE. 2010. Mechanisms involved in the developmental programming of adulthood disease. *Biochem J* **427**: 333–347.
- Waterland RA, Kellermayer R, Laritsky E, Ravco-Soion P, Harris RA, Travisano M, Zhang W, Torskaya MS, Zhang J, Shen L, et al. 2010. Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet* **6**: e1001252. doi: 10.1371/journal.gen.1001252.
- Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schübeler D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37**: 853–862.

Received November 14, 2011; accepted in revised form June 8, 2012.