

HUMAN GENETICS

Establishment of environmentally sensitive DNA methylation states in the very early human embryo

Noah J. Kessler¹, Robert A. Waterland², Andrew M. Prentice¹, Matt J. Silver^{1*}

The molecular mechanisms responsible for the developmental origins of later disease are currently unknown. We previously demonstrated that women's periconceptional nutrition predicts their offspring's DNA methylation at metastable epialleles (MEs). We present a genome-wide screen yielding 687 MEs and track their trajectories across nine developmental stages in human in vitro fertilization embryos. MEs exhibit highly unusual methylation dynamics across the implantation-gastrulation transition, producing a large excess of intermediate methylation states, suggesting the potential for differential programming in response to external signals. Using a natural experiment in rural Gambia, we show that genomic regions sensitive to season of conception are highly enriched for MEs and show similar atypical methylation patterns. MEs are enriched for proximal enhancers and transcription start sites and are influenced by genotype. Together, these observations position MEs as distinctive epigenomic features programmed in the early embryo, sensitive to genetic and periconceptional environment, and with the potential to influence phenotype.

INTRODUCTION

DNA methylation is a widely studied epigenetic mark that plays a key role in the transcriptional regulation of a number of cellular processes in mammals including cell differentiation, genomic imprinting, and X-inactivation (1). Early embryonic development represents a critical window for the establishment of the methylome, when the human preimplantation embryo undergoes substantial remodeling, with widespread erasure of gametic methylation marks as the embryo transitions to a pluripotent state (1–3). Animal and human studies indicate that the methylation state can be influenced by the environment of the early embryo, suggesting a potential role for methylation and related epigenetic marks in mediating the effects of early-life nutritional and other environmental stressors on later health and disease (4–6).

Metastable epialleles (MEs) are genomic regions that show systemic (cross-tissue) interindividual variation in methylation, indicating establishment of the variable methylation state in the preimplantation embryo, before gastrulation (7). Murine and human MEs have been associated with the presence of neighboring transposable elements and are influenced by nutritional and other environmental stressors at periconception (8–13). Variable methylation at MEs was originally defined as occurring in the absence of genetic variation (7), and MEs have been widely studied in isogenic mice (14–16), demonstrating that systemic, stochastic variation in methylation state can occur independent of genotype. As we expand the search for ME regions into genetically heterogeneous human populations, we suggest that this definition should be extended to include genomic regions whose epigenetic state is under partial but nondeterministic genetic influence.

¹Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. ²U.S. Department of Agriculture/Agricultural Research Service Children's Nutrition Research Center, Departments of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

*Corresponding author. Email: matt.silver@lshtm.ac.uk

RESULTS

Screen for human MEs

We previously reported the first-in-human genome-wide screen for MEs using whole-genome bisulfite-seq (WGBS) data from two tissues in two North American Caucasian individuals (8). This screen was limited by its inclusion of only two germ layer lineages (mesoderm and ectoderm) and its small sample size. For the current analysis, we produced an updated list of putative human ME regions by extending the previous ME screen to include samples from three new individuals and from an endodermal tissue (17). These 687 ME regions exhibit systemic interindividual variation in DNA methylation that is distinct from the patterns in 5902 genomic clustered “control” regions generated from the whole-genome background using the same clustering parameters used to find the ME regions (Fig. 1, table S1, fig. S1A, and Materials and Methods).

Chromatin state at MEs was assessed using the histone-based 15-state chromHMM model (18) in three adult tissues (one derived from each germ layer), generated as part of the Roadmap Epigenomics Project (19). We observed that ME regions are more likely to be associated with predicted enhancers, transcription start sites, and zinc finger genes, and less likely to be associated with transcribed regions or regions with low levels of histone marks, compared to control regions and to genomic background in all three tissues surveyed (Fig. 2A). Furthermore, MEs are more likely to be near certain classes of transposable elements than would be expected given the genomic distribution of CG dinucleotides (CpGs), with significant enrichment for proximity to endogenous retroviral sequence 1 (ERV1) and endogenous retrovirus group K (ERVVK) elements [$P = 4.7 \times 10^{-8}$, $P = 1.4 \times 10^{-14}$, Fisher's exact test (FET); see Materials and Methods and table S2]. This is in line with our previous observations (8) and with observations in mice (15).

Methylation dynamics of MEs during embryonic development

Patterns of systemic interindividual variation at MEs (Fig. 1) suggest that the methylation state is established in the early embryo, prior to separation into germ layers at gastrulation (7), but this has not yet been demonstrated in human embryonic samples. We analyzed human methylomes from early developmental stages in Chinese

Copyright © 2018
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
License 4.0 (CC BY).

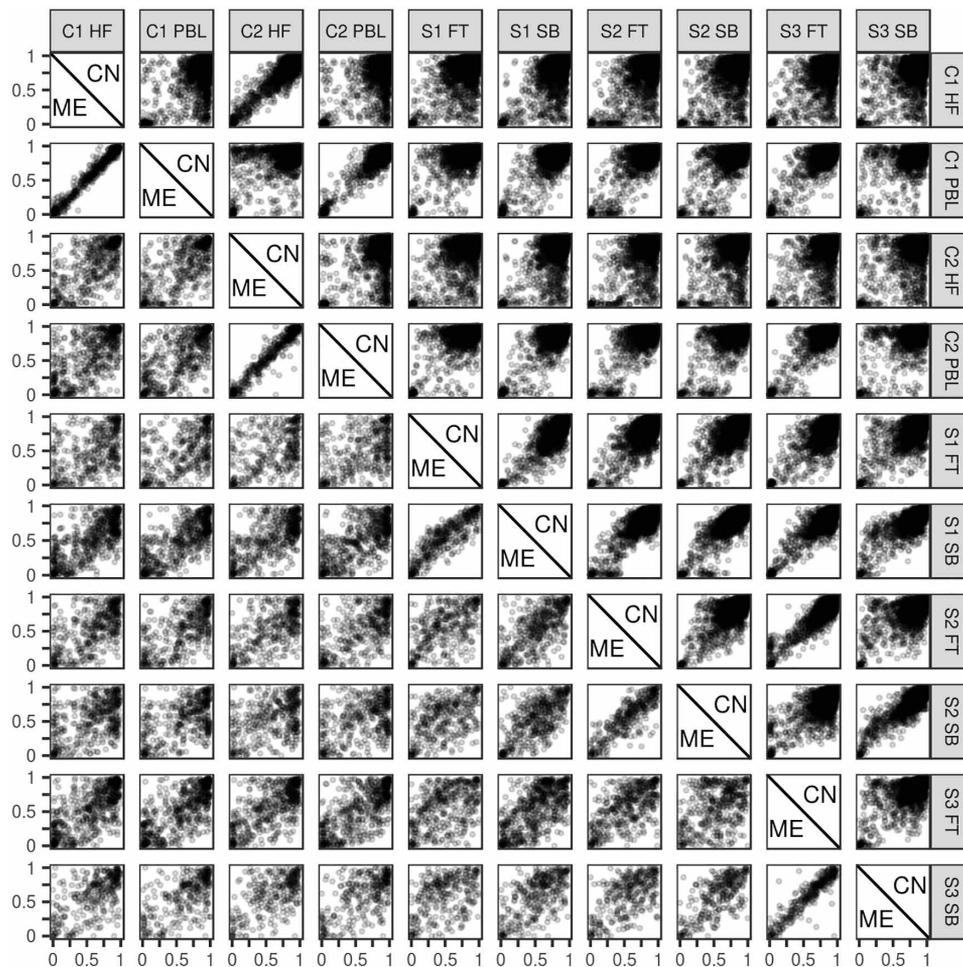


Fig. 1. Methylation of ME and control regions in samples used for the ME screen. Methylation distributions in 687 ME and 5902 control regions. X axes represent methylation in the sample given by the label at the top of the column. Y axes, the same, but given by label at the end of the row. Each point represents mean methylation across all CpGs in a single region. Samples are listed by individual and tissue. HF, hair follicle; PBL, peripheral blood lymphocytes; FT, fat; SB, small bowel. Methylation in MEs is on the bottom (left) of the diagonal; methylation of controls (C) is on the top (right).

embryos using public data (20). This data set consists of reduced representation bisulfite-seq (RRBS) data from two to four biological replicates at each of nine developmental stages, from gametes, through cleavage-stage embryos, to blastocyst and differentiated embryonic tissue (table S3). More than 3.5 million CpG sites were covered in at least one of the replicates (fig. S1B), and despite RRBS covering only approximately 10% of genome-wide CpGs, a substantial proportion (44%) of our ME regions were covered in the Guo *et al.* data set (table S4).

As a first step, we replicated the analysis from Guo *et al.* (20) showing genome-wide changes in mean methylation at each developmental stage (Fig. 3A). Expected patterns of genome-wide demethylation and remethylation at periconception and across the gastrulation transition are visible in this data set. In contrast to genomic background, methylation at MEs is consistently lower at all developmental stages, particularly in sperm (Fig. 3A). Mean methylation in clustered control regions is higher than in genomic background, making the contrast with ME methylation even greater (fig. S2). Similar patterns are evident in a mostly single cell-derived data set from the same group (fig. S3 and Materials and Methods) (21).

Since change in global mean methylation is a relatively crude measure of methylation dynamics in the early embryo, we next considered the distribution of methylation at each stage (Fig. 3B). As suggested by Fig. 3A, we observed an increased proportion of low methylation states in ME regions relative to background (that is, all CpGs covered in the data set).

Most striking, however, was a marked increase in intermediate methylation states (sites with 10 to 90% methylation) within ME regions in the post-implantation embryo (“embryonic liver”) (Fig. 3B). Here, 52.3% of CpGs in ME regions show intermediate methylation versus 20.1% in genomic background and 26.6% in control regions ($P < 2.2 \times 10^{-16}$, chi-squared test for difference in proportions for both comparisons; fig. S5B). Genome-wide, there was a tendency toward a bimodal distribution of very high or low methylation after gastrulation (Fig. 4, A and B, and fig. S4). In contrast, in ME regions, a substantial proportion of CpGs are intermediately methylated in post-gastrulation tissue, irrespective of starting methylation state in (pre-gastrulation) inner cell mass (ICM) tissue. These distinctive patterns are evident in other fetal tissues (intestine, kidney, and lung) obtained from the Roadmap Epigenomics Project

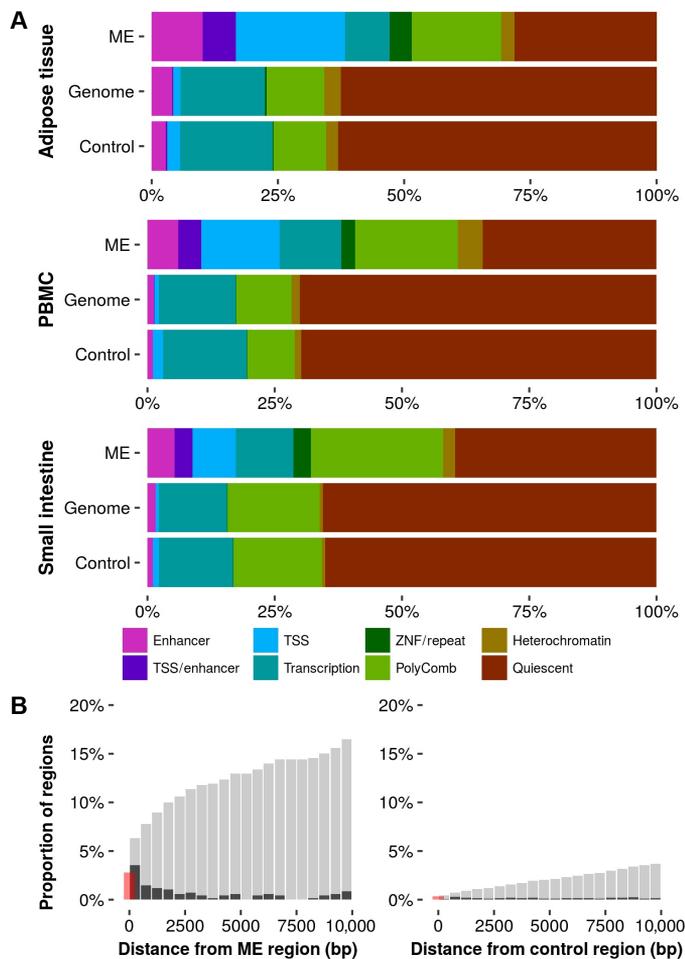


Fig. 2. Characteristics of ME and control regions. (A) Proportion of MEs, control regions, and the whole genome, overlapping predicted chromatin states from three Roadmap samples (one each from endoderm, mesoderm, and ectoderm). Fifteen chromHMM states have been reduced to eight after combining states, which represent similar genomic features. Quiescent loci are those with low signal in all of the histone marks used as input to the chromHMM algorithm; ZNF regions are those overlapping zinc finger protein–encoding genes. (B) Proximity of MEs and controls to ZFP57 binding sites. Black bars show the proportion of regions with a binding site within the distance marked on the x axis. Gray bars show the cumulative proportion of regions within a given distance of a binding site. Red bars show the proportion of regions that overlap a binding site. PBMC, peripheral blood mononuclear cell; TSS, transcription start site.

(19) and are also evident when comparing MEs with clustered control regions (fig. S5, A and B).

Intermediate methylation states in pooled samples may result from interindividual variation with low intraindividual variation, heterogeneity of cell methylation states within individuals, or a combination of both. In the Guo *et al.* data set, most replicates at each developmental stage consist of cells derived from several embryos (table S3). However, each embryonic liver replicate represents only one conceptus, allowing us to assess the methylation profile of three individuals separately. In each of these replicates, we still observe a large proportion of intermediate methylation states in ME regions (Fig. 4C), indicating that intermediate methylation is driven by intratissue heterogeneity of methylation states within an

individual, rather than by interindividual differences. A read-level analysis suggests that intermediate methylation at the CpG level arises from a combination of partially (10 to 90%) and fully methylated/unmethylated (molecule-specific) reads, within both ME and clustered control regions (Fig. 5A).

To test this possibility, we devised a read homogeneity index (RHI) to evaluate the extent to which intermediately methylated clusters are driven by mixtures of reads each with high methylation homogeneity. This normalized index accounts for differences due to cluster size and average methylation (see Materials and Methods). Clusters containing more reads with long runs of methylated or unmethylated CpGs will have a higher RHI (see Fig. 5C for a representative illustration). We found that intermediately methylated ME clusters have a significantly higher RHI than equivalent control clusters (Mann-Whitney, $P < 3 \times 10^{-5}$; Fig. 5B), indicating increased levels of molecule-specific methylation at MEs.

Association of genetic variation and proximal protein binding sites with ME methylation

The above observations in single replicates indicate that cellular differences in methylation state occur independent of genetic variation. It remains the case, however, that methylation at MEs might be influenced by genotype (22). Our ME screen filtered out differences in methylation for CpGs within 60 base pairs (bp) of a genetic variant (see Materials and Methods), but observed interindividual differences may be driven by more distant genetic variation. The identification of mQTL—genetic variants influencing CpG methylation—requires large sample sizes, and so far, sufficiently powered studies have only been conducted using methylation arrays that assay a small fraction of the methylome. We therefore assessed potential genetic influence on DNA methylation at MEs identified in our WGBS screen by analyzing mQTL identified in a large UK study (see Materials and Methods) (23). This analysis identified genetic influences from 8.3 million common single-nucleotide polymorphisms (SNPs) on approximately 400,000 CpGs assayed in 3948 blood samples taken at five time points. We observed highly significant enrichment for mQTL at CpGs in ME regions [odds ratio (OR), 6.2; $P < 10^{-10}$, FET] but not in control regions (OR, 1.3; $P = 0.09$). Estimates of the proportion of methylation variance explained by cis and trans mQTL at MEs indicate that mQTL explain only a small to moderate proportion of methylation variance at mQTL-associated ME-CpGs {median proportion of variance explained, 0.23 [interquartile range (IQR), 0.057 to 0.46] in birth samples and 0.24 (IQR, 0.055 to 0.49) in adult samples; Fig. 3D and Materials and Methods}, suggesting that stochastic and/or environmental factors play a major role in driving the distinctive early embryo dynamics observed at MEs.

Control of methylation dynamics in the preimplantation embryo is critical in genomic imprinting (24). The Krüppel associated box (KRAB) zinc finger protein ZFP57 and its binding partner tripartite motif-containing 28 (TRIM28), along with the multi-zinc finger protein CCCTC-binding factor (CTCF), have been implicated in the maintenance of parental methylation marks in the early embryo (25–27). We therefore speculated that these might influence methylation dynamics at MEs. Using public human chromatin immunoprecipitation sequencing (ChIP-seq) data, we found significant enrichment for proximal ZFP57, CTCF, and TRIM28 binding sites in ME regions compared to genomic background (respectively $P < 2.2 \times 10^{-16}$, $P < 2.2 \times 10^{-16}$, $P = 4.1 \times 10^{-7}$, FET; table S2

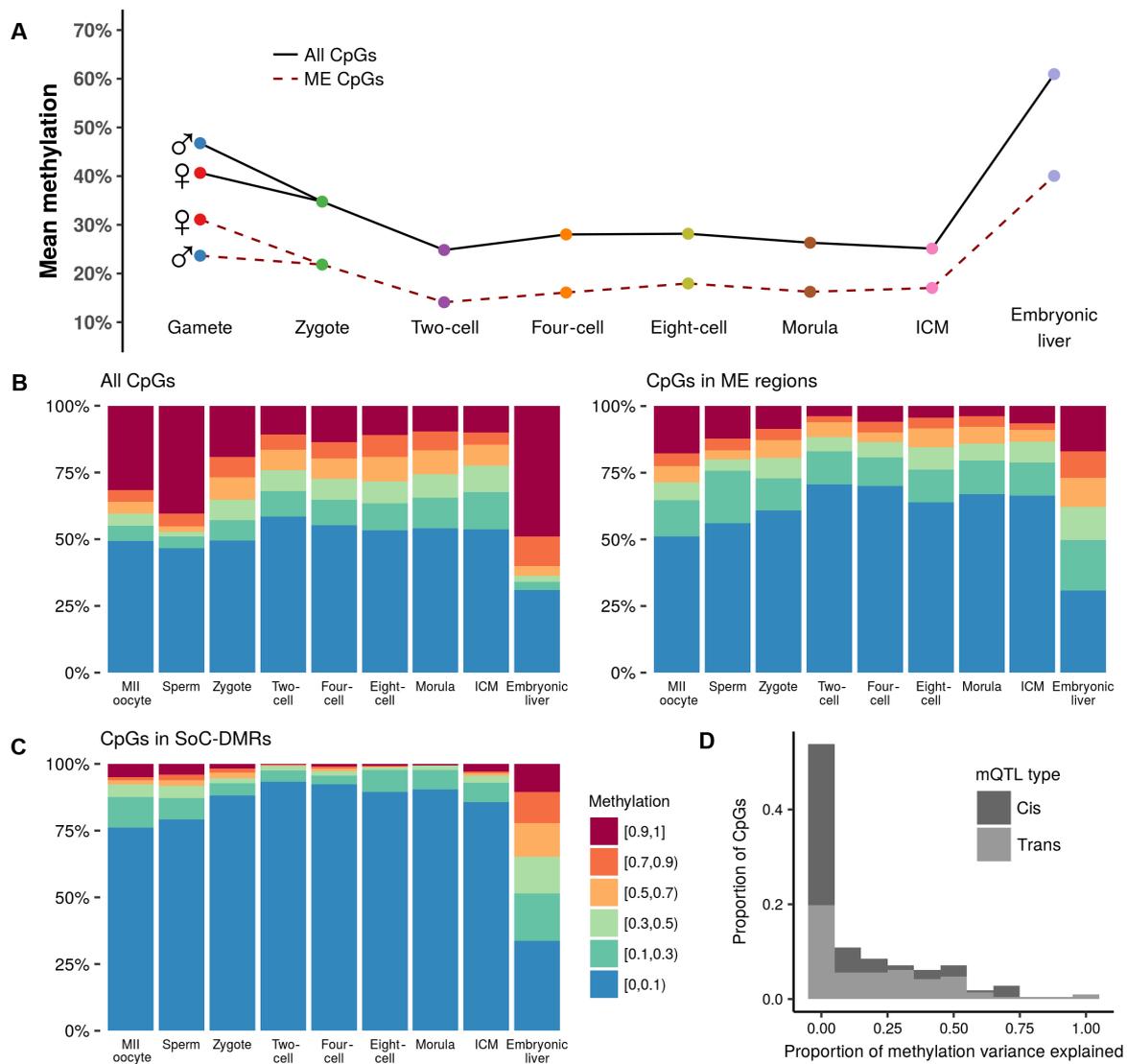


Fig. 3. Methylation at MEs, all-RRBS background, and SoC-associated loci, and potential genetic influences. Within each developmental stage in all plots, CpGs are counted once for each replicate for which there was sufficient read depth. (A) Mean methylation at each developmental stage assayed by Guo *et al.* Solid line: mean methylation in all-RRBS background ($n = 3,679,155$ CpGs); dashed line: mean methylation at ME regions covered by RRBS ($n = 302$ regions; 2098 CpGs). (B) Distribution of methylation across developmental stages in all-RRBS background CpGs (left) and at CpGs within ME regions (right). See legend in (C). (C) Distribution of methylation at each stage in previously identified season of conception-associated differentially methylated regions (SoC-DMRs; see Materials and Methods). (D) Proportion of methylation variance explained by genetic variants in cis and in trans at ME-CpGs. These data are obtained from a previous comprehensive screen for methylation quantitative loci (mQTL)-CpGs in a large European study (23).

and Materials and Methods). For example, 16.3% of ME regions are within 10 kb of a ZFP57 binding site, compared with 3.7% of clustered control regions (7.8% versus 0.7% using a 1-kb distance threshold; 2.8% versus 0.3% using exact overlaps; Fig. 2B and table S1).

To investigate the potential influence of ZFP57 binding on ME methylation, we analyzed data from a study of individuals homozygous for ZFP57 mutations associated with transient neonatal diabetes, a rare imprinting disorder (28). By comparing individuals with ZFP57 mutations with matched controls, this study found 61 DMRs (ZFP57m-DMRs), which were common to at least two patients with a homozygous mutation. Despite the small number of ZFP57m-DMRs, five overlapped an ME region and a further two were within

10 kb (table S5). All were hypomethylated in the individuals with the mutation. In contrast, no control clusters overlap or are within 10 kb of a ZFP57m-DMR.

Influence of periconceptual environment on ME methylation

We previously reported associations between season of conception (SoC) and DNA methylation at a small number of MEs in a Gambian population (8, 29, 30). In this rural community in sub-Saharan Africa, seasonality is linked to significant differences in maternal diet and circulating maternal methyl donor biomarkers (31), suggesting that, as observed at murine MEs, human MEs may be sensitive to the periconceptual nutritional environment (8, 11, 30). We therefore hypothesized that Gambian SoC-associated genomic regions

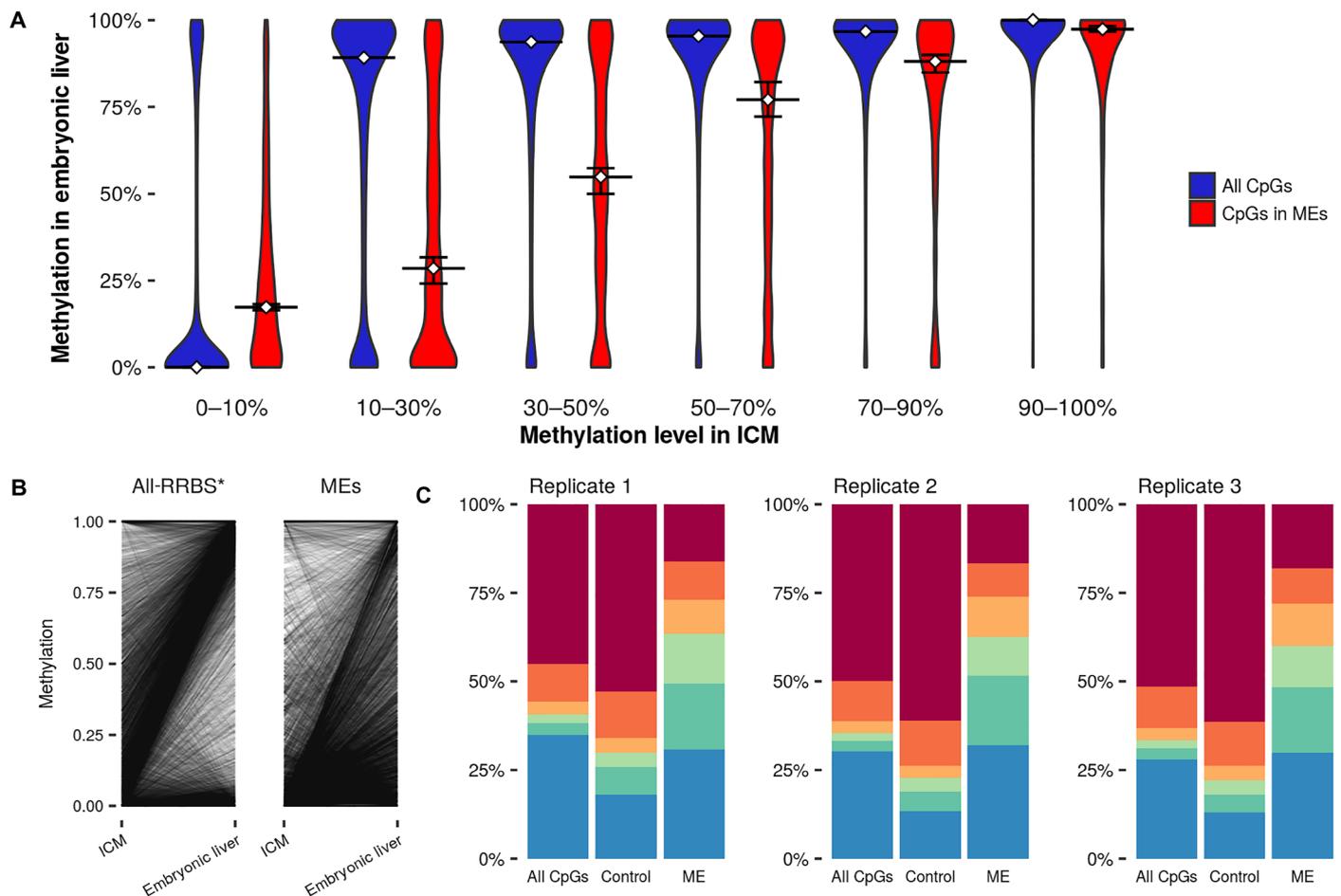


Fig. 4. Methylation dynamics around the gastrulation transition. (A) Distribution of CpG methylation in embryonic liver, stratified by methylation level in ICM. Horizontal lines show the median methylation in embryonic liver. Error bars represent 95% bootstrapped confidence intervals. (B) Change in methylation at individual CpGs across the transition. Each line represents the change in methylation at a CpG for a single combination of replicates from each stage. *Because of the large number of data points, "all-RRBS" plot shows a random 0.1% sample of all possible ICM-embryonic liver transitions. (C) Distribution of methylation states in each embryonic liver replicate.

may show patterns of early embryo methylation dynamics similar to those observed at MEs from our WGBS screen in North American Caucasians.

To investigate this possibility, we curated a list of 21 SoC-DMRs previously identified in epigenome-wide association studies from two independent Gambian studies (see Materials and Methods). Gambian SoC-DMRs show extremely low levels of methylation in the Chinese gametes and pre-gastrulation embryonic tissues assayed by Guo *et al.*, with strong enrichment of intermediate methylation states relative to background in the post-gastrulation embryo ($n = 17$ SoC-DMR regions with ≥ 1 CpG in the Guo *et al.* data set; Fig. 3C), reminiscent of patterns observed at MEs (Fig. 3B). SoC-DMRs were highly enriched for overlapping MEs (5 of 21 DMRs; $P < 2.2 \times 10^{-16}$, FET) and for mQTL ($P = 7.1 \times 10^{-16}$, FET).

DISCUSSION

We have presented the first-in-human characterization of methylation dynamics at MEs in early embryos. Our data suggest that the gastrulation transition is key to the establishment of intermediate methylation states at MEs and that these states are driven by intra-

tissue variegation effects, as originally proposed (7). In addition, our analysis suggests that interindividual variation at MEs is influenced by at least three factors: stochastic or probabilistic processes, periconceptual environmental exposures, and genomic context. The last of these is notable since previous studies of murine MEs have generally considered isogenic populations, so epigenetic metastability in humans was assumed to be definitively free of genetic effects. We note, however, that SNP heritability estimates from array-based mQTL studies necessarily cover relatively few ME-CpGs in our WGBS screen, so further work is required to better understand the influence of genotype on ME methylation.

Our work confirms previous findings in humans and in mice of a link between metastability and proximal transposable elements and further indicates a potential role for zinc finger proteins including CTCF and ZFP57. Recent work characterizing haplotype-dependent allele-specific methylation (hap-ASM) in humans has identified polymorphic CTCF binding sites as important drivers of hap-ASM including at ZFP57, indicating that CpG methylation at this locus is at least partially dependent on haplotype (32). Van Baak *et al.* (33) recently observed interindividual variation of methylation between individuals sharing the same haplotype at ZFP57,

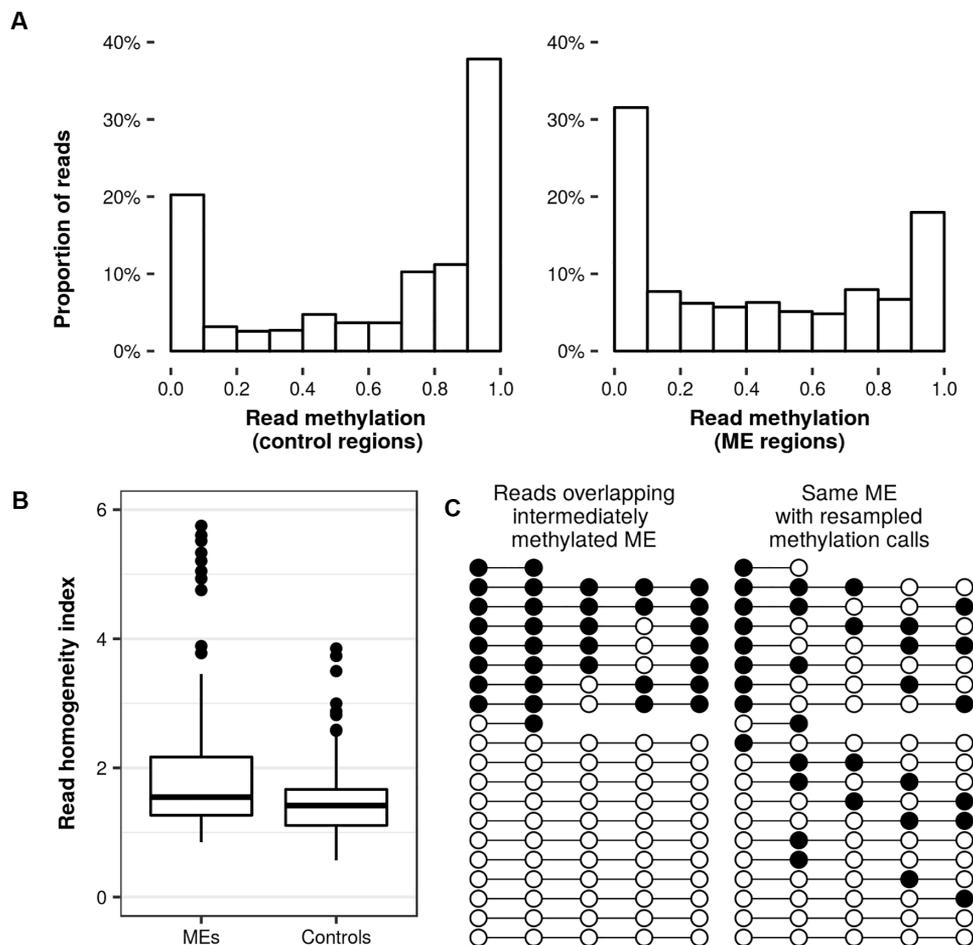


Fig. 5. Read-level methylation analysis. (A) Distribution of read-level methylation for reads overlapping CpGs with intermediate methylation (10 to 90%) in clustered control regions (left, 31.8K reads) and ME regions (right, 22.5K reads). (B) RHI scores for MEs ($n = 133$) and control regions ($n = 200$) covered in the Guo *et al.* data show a significantly higher average RHI in MEs (Mann-Whitney, $P < 5 \times 10^{-5}$; nine MEs and two control outliers with RHI > 6 not shown). (C) Lollipop plots showing methylation calls overlapping an ME (left, reads ordered by read-level methylation) and a random resampling of the methylation calls, preserving the location of the reads (right). At this ME, the observed methylation calls result in a transition count of 11, while the resampled example has a transition count of 34, which is the median of 1000 resamplings performed on this region. The RHI of this region is thus $34/11$ or 3.1.

indicating a potential role for stochastic processes and environmental influence in the context of genetic effects, as we observe at MEs in embryonic liver tissue from single embryos. As MEs are frequently located near ZFP57 binding sites, altered expression of ZFP57 due to variable methylation near its transcription start site could contribute to the variance in methylation at MEs. This notion is supported by our finding that several ZFP57 binding site-proximal MEs are among DMRs associated with human ZFP57 mutations.

Chromatin state analysis suggests that, compared to background, MEs tend to be located within enhancers and proximal to transcription start sites, indicating potential effects on transcriptional regulation. While further work is required to link methylation status to gene expression, this analysis positions MEs as possible mediators of phenotypic plasticity in response to periconceptual exposures, as has been proposed in previous models of developmental programming (34). The presence of variably methylated sites that are linked to gene function, and are susceptible to environmental and genetic influence, has been proposed as a potential adaptive mechanism (35), positioning MEs as prime candidates for investigating adaptive responses to changing environments.

MATERIALS AND METHODS

ME screen using WGBS data

Two WGBS data sets were used to search for human MEs: that used in our previous screen (two Caucasian adult males, two tissues: peripheral blood and hair follicle) (8) and an additional data set with three individuals in the Roadmap Epigenomics Project (S1: male, 3 years old, Caucasian/African-American; S2: female, 30 years old, Caucasian; S3: male, 34 years old, Caucasian; two tissues: small bowel and fat) (19). Bismark v0.14 (36) was used to map the raw reads and extract methylation at CpGs. In each sample, CpGs with less than 10 times the read depth were filtered out. ME regions were generated using a three-step process: (i) identify individual CpGs showing the hallmarks of metastability (“ME-CpGs”) in each data set, (ii) run a clustering algorithm on these loci, and (iii) filter out clusters that contain many non-ME-CpGs. For the first step, CpGs were required to have at least a 15% absolute difference in methylation between individuals and an absolute intertissue methylation difference no greater than one-third of the absolute interindividual difference. This 15% methylation difference was assessed separately for the two WGBS data sets. Within each of these two data sets,

SNPs were called in each individual from the bisulfite-converted reads using bis-SNP (37). As a precaution against overselecting for methylation variation driven by proximal variants (for example, CpGs with ASM), any CpG within 60 bp of a SNP (22) was excluded from consideration in the data set containing the SNP. ME clusters were generated by filtering out ME-CpGs that were not within 300 bp of another ME-CpG, and then requiring that each resulting cluster had four or more ME-CpGs. The final list of 687 ME regions was then filtered from those clusters by selecting only those which had at least twice as many ME-CpGs as non-ME-CpGs.

Control cluster generation

We first considered deriving a set of control regions as those showing interindividual variation without intertissue concordance; however, few such regions exist in our data sets. We therefore chose regions by applying clustering parameters identical to those used for ME clusters to all genomic CpGs covered in our WGBS data sets, irrespective of methylation. A sample of the resulting regions, matching the joint distributions of region size and number of CpGs per region found in the MEs, was then used to generate 5902 control regions (fig. S1A).

Enrichment for proximal genomic features

Transposable element regions [LINE (long interspersed nuclear elements), SINE (short interspersed nuclear elements), and LTR (long terminal repeats)] of the human genome (hg19), determined by RepeatMasker, were downloaded from the UCSC (University of California, Santa Cruz) (hg19 annotations repository. Human cell line ChIP-seq binding sites were downloaded from the GEO (Gene Expression Omnibus) or ENCODE (Encyclopedia of DNA Elements) [ZFP57: human embryonic kidney 293T (HEK293T) cells, GSM2466450; TRIM28: HEK293T cells, ENCFF002CRN; CTCF: H1-hESC (human embryonic stem cells), ENCFF002CDS]. For enrichment testing, the genome was divided into 1000-bp “bins” or “tiles” (8, 20), and the number of tiles overlapping a feature of interest was compared with the number of tiles overlapping features within 10 kb of an ME. FET was then used to determine the OR and *P* value for enrichment of ME-proximal features. The list of ZFP57-mutant DMRs was from Bak *et al.* (28) (table S5).

Guo *et al.* and Zhu *et al.* data analysis

RRBS methylation data from all replicates of each of the stages analyzed were downloaded from GEO (accession number GSE49828). Sites were considered covered if total read depth from both strands was at least 10. CpG coverage within each sample and overall is summarized in fig. S1B and table S3. For the read-level analyses in the embryonic liver samples, raw reads were downloaded from the Sequence Read Archive (SRA). Reads were trimmed and filtered for quality and size by Trim Galore! and mapped by Bismark v0.18 (36), without deduplication, as recommended by the Bismark authors for RRBS data. Reads were further filtered to exclude those containing fewer than four CpGs. Read-level methylation was then determined by taking the average of all CpGs included in the read (that is, the number of methylated CpGs divided by the total number of CpGs present on the read). Only CpGs with 10 times the coverage were considered when computing CpG-level methylation. CpGs falling outside the ME or control regions were not analyzed, and CpGs with >200 times the coverage were ignored to prevent highly overrepresented regions from dominating the analysis. Bootstrapped confi-

dence intervals for medians in Fig. 3A were generated by resampling the original data 1000 times.

Data from Zhu *et al.* (21) were downloaded from GEO (accession number GSE81233), including all single-cell and bulk cell data sets from MII oocyte, sperm, zygote, two-cell, four-cell, eight-cell, morula, ICM, and each of the three embryonic heart samples. All data were gathered by developmental time point, and all CpGs with read depth ≥ 10 were included in the plots of developmental time point-level means. Because of the small number of samples at many developmental time points (compounded by the fact that single-cell reads only provide two molecules' worth of information), no further analysis on these samples was performed.

Read homogeneity index

For each ME and control cluster, all reads from the Guo *et al.* embryonic liver replicates overlapping the cluster were combined, and any CpGs outside of the cluster were ignored. We initially considered using a published method for this, such as methylation haplotype load (38), but developed the RHI instead as other methods assume a large number of reads overlapping a fixed set of adjacent CpGs, whereas the bisulfite-seq reads from Guo *et al.* sometimes contain uncalled cytosines (for example, due to SNPs or a low-quality base call) or only overlap the first or last one to two CpGs of an ME/control region (for example, the short reads in Fig. 5C). To account for the effects of differences in read and cluster size, and overall methylation within a cluster, a normalized RHI was calculated as follows. First, the observed number of methylation “transitions” was counted, a transition being any instance where two adjacent CpGs have different methylation calls. Next, the same measure was computed, this time with an equal number of methylated CpGs randomly distributed across reads within the cluster. This randomization was repeated 1000 times. Finally, the RHI was calculated as the median transition count across all 1000 randomizations, divided by the observed (empirical) transition count. The RHI thus provides a measure of the degree to which reads are consistently methylated or unmethylated, compared to reads where methylated CpGs are randomly distributed. Note that the RHI will generally be greater than 1 due to the tendency for methylation states at neighboring CpGs to be correlated.

mQTL analysis

CpGs associated with mQTL on the Illumina Infinium Human-Methylation450 BeadChip were downloaded from www.mqtl.org. These were identified using samples from children (cord blood, $n = 771$; childhood, $n = 834$; adolescence, $n = 837$) and their mothers (pregnancy, $n = 764$; middle age, $n = 742$). See Gaunt *et al.* for further details (23). For the current analysis, MEs were tested for enrichment of mQTL-associated CpGs at any time point, relative to “reliable” non-mQTL-associated CpGs on the 450K array [see Gaunt *et al.* (23)], using FET. Data on proportion of methylation variance explained by cis (± 1 Mb of CpG probe) and trans (non-cis) mQTL (“SNP heritability” in Gaunt *et al.*) were obtained from one of the authors (see Acknowledgments). Figure 3D shows the distribution of variance explained for CpGs overlapping MEs in “middle age” samples, this being the closest time point to the samples that were used for the ME screen.

Gambian SoC-DMRs

SoC-DMRs were identified by taking the top 10 ranking DMRs [by family-wise error rate (FWER) using the “bumphunting” method (39)]

from two SoC-association analyses in independent samples conducted using the Illumina 450K array in (i) PBL from Gambian infants (8) and (ii) PBL from Gambian 2-year-olds (40). A further seven SoC-DMRs ranked in the top 50 (by FWER) in both data sets were also included. This resulted in a total of 21 distinct (nonoverlapping) SoC-DMRs. Information on ethical approvals for each of the data sets analyzed in this study can be found in the source publications cited above.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/7/eaat2624/DC1>

- Fig. S1. ME and control region sizes, and Guo *et al.* methylation coverage.
 Fig. S2. Mean methylation at MEs and clustered control regions assayed by Guo *et al.*
 Fig. S3. Mean methylation at all CpGs and at MEs and clustered control regions assayed by Zhu *et al.* (21).
 Fig. S4. Methylation dynamics at the ICM-to-embryonic liver transition.
 Fig. S5. ME background comparisons in other fetal tissues, and methylation in control clusters.
 Table S1. MEs identified in genome-wide screen.
 Table S2. Enrichment of proximal genomic features in MEs.
 Table S3. Number of CpGs covered in each replicate of RRBS data from Guo *et al.*
 Table S4. Size of ME and control regions, and their coverage in Guo *et al.* RRBS data.
 Table S5. Overlap of Bak *et al.* (28) ZFP57-mutant DMRs with MEs.

REFERENCES AND NOTES

- H. Okae, H. Chiba, H. Hiura, H. Hamada, A. Sato, T. Utsunomiya, H. Kikuchi, H. Yoshida, A. Tanaka, M. Suyama, T. Arima, Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet.* **10**, e1004868 (2014).
- H. J. Lee, T. A. Hore, W. Reik, Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell* **14**, 710–719 (2014).
- Z. D. Smith, M. M. Chan, K. C. Humm, R. Karnik, S. Mekhoubad, A. Regev, K. Eggan, A. Meissner, DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
- P. D. Gluckman, M. A. Hanson, C. Cooper, K. L. Thornburg, Effect of in utero and early-life conditions and adult health and disease. *N. Engl. J. Med.* **359**, 1523–1524 (2008).
- R. A. Waterland, C. Garza, Potential mechanisms of metabolic imprinting that lead to chronic disease. *Am. J. Clin. Nutr.* **69**, 179–197 (1999).
- E. W. Tobj, R. C. Sliker, A. D. Stein, H. E. D. Suchiman, P. E. Slagboom, E. W. van Zwet, B. T. Heijmans, L. H. Lumey, Early gestation as the critical time-window for changes in the prenatal environment to affect the adult human blood methylome. *Int. J. Epidemiol.* **44**, 1211–1223 (2015).
- V. K. Rakan, M. E. Blewitt, R. Druker, J. I. Preis, E. Whitelaw, Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002).
- M. J. Silver, N. J. Kessler, B. J. Hennig, P. Dominguez-Salas, E. Laritsky, M. S. Baker, C. Coarfa, H. Hernandez-Vargas, J. M. Castilino, M. N. Routledge, Y. Y. Gong, Z. Herceg, Y. S. Lee, K. Lee, S. E. Moore, A. J. Fulford, A. M. Prentice, R. A. Waterland, Independent genomewide screens identify the tumor suppressor *VTRNA2-1* as a human epiallele responsive to periconceptional environment. *Genome Biol.* **16**, 118 (2015).
- P. Kuehnen, M. Mischke, S. Wiegand, C. Sers, B. Horsthemke, S. Lau, T. Keil, Y.-A. Lee, A. Grueters, H. Krude, An Alu element-associated hypermethylation variant of the *POMC* gene is associated with childhood obesity. *PLoS Genet.* **8**, e1002543 (2012).
- S. Finer, M. S. Iqbal, R. Lowe, B. W. Ogunkolade, S. Pervin, C. Mathews, M. Smart, D. S. Alam, G. A. Hitman, Is famine exposure during developmental life in rural Bangladesh associated with a metabolic and epigenetic signature in young adulthood? A historical cohort study. *BMJ Open.* **6**, e011768 (2016).
- R. A. Waterland, R. L. Jirtle, Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell Biol.* **23**, 5293–5300 (2003).
- A. J. Bernal, R. L. Jirtle, Epigenomic disruption: The effects of early developmental exposures. *Birth Defects Res. A Clin. Mol. Teratol.* **88**, 938–944 (2010).
- C. Faulk, J. H. Kim, O. S. Anderson, M. S. Nahar, T. R. Jones, M. A. Sartor, D. C. Dolinoya, Detection of differential DNA methylation in repetitive DNA of mice and humans perinatally exposed to bisphenol A. *Epigenetics* **11**, 489–500 (2016).
- H. D. Morgan, H. G. E. Sutherland, D. I. K. Martin, E. Whitelaw, Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23**, 314–318 (1999).
- V. K. Rakan, S. Chong, M. E. Champ, P. C. Cuthbert, H. D. Morgan, K. V. K. Luu, E. Whitelaw, Transgenerational inheritance of epigenetic states at the murine *Axin^{Fu}* allele occurs after maternal and paternal transmission. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2538–2543 (2003).
- H. Oey, L. Isabel, P. Hickey, B. Ebad, E. Whitelaw, Genetic and epigenetic variation among inbred mouse littermates: Identification of inter-individual differentially methylated regions. *Epigenetics Chromatin* **8**, 54 (2015).
- M. D. Schultz, Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urlich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang, J. R. Ecker, Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
- J. Ernst, M. Kellis, ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
- A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfennig, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthal, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- H. Guo, P. Zhu, L. Yan, R. Li, B. Hu, Y. Lian, J. Yan, X. Ren, S. Lin, J. Li, X. Jin, X. Shi, P. Liu, X. Wang, W. Wang, Y. Wei, X. Li, F. Guo, X. Wu, X. Fan, J. Yong, L. Wen, S. X. Xie, F. Tang, J. Qiao, The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).
- P. Zhu, H. Guo, Y. Ren, Y. Hou, J. Dong, R. Li, Y. Lian, X. Fan, B. Hu, Y. Gao, X. Wang, Y. Wei, P. Liu, J. Yan, X. Ren, P. Yuan, Y. Yuan, Z. Yan, L. Wen, L. Yan, J. Qiao, F. Tang, Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet.* **50**, 12–19 (2018).
- A. L. Teh, H. Pan, L. Chen, M.-L. Ong, S. K. Dogra, J. Wong, J. L. MacIsaac, S. M. Mah, L. M. McEwen, S.-M. Saw, K. M. Godfrey, Y.-S. Chong, K. Kwek, C.-K. Kwok, S.-E. Soh, M. F. F. Chong, S. Barton, N. Karnani, C. Y. Cheong, J. P. Buschdorf, W. Stunkel, M. S. Kobor, M. J. Meaney, P. D. Gluckman, J. D. Holbrook, The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res.* **24**, 1064–1074 (2014).
- T. R. Gaunt, H. A. Shihab, G. Hemani, J. L. Min, G. Woodward, O. Lyttleton, J. Zheng, A. Duggirala, W. L. McArdle, K. Ho, S. M. Ring, D. M. Evans, G. D. Smith, C. L. Relton, Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* **17**, 61 (2016).
- A. C. Ferguson-Smith, Genomic imprinting: The emergence of an epigenetic paradigm. *Nat. Rev. Genet.* **12**, 565–575 (2011).
- N. Takahashi, D. Gray, R. Strogantsev, A. Noon, C. Delahaye, W. C. Skarnes, P. H. Tate, A. C. Ferguson-Smith, ZFP57 and the targeted maintenance of postfertilization genomic imprints. *Cold Spring Harb. Symp. Quant. Biol.* **80**, 177–187 (2015).
- S. Quenneville, G. Verde, A. Corsinotti, A. Kapopoulou, J. Jakobsson, S. Offner, I. Baglivo, P. V. Pedone, G. Grimaldi, A. Riccio, D. Trono, In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* **44**, 361–372 (2011).
- R. Strogantsev, F. Krueger, K. Yamazawa, H. Shi, P. Gould, M. Goldman-Roberts, K. McEwen, B. Sun, R. Pedersen, A. C. Ferguson-Smith, Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* **16**, 112 (2015).
- M. Bak, S. E. Boonen, C. Dahl, J. M. D. Hahnemann, D. J. D. G. Mackay, Z. Tümer, K. Granskov, I. K. Temple, P. Guldborg, N. Tommerup, Genome-wide DNA methylation analysis of transient neonatal diabetes type 1 patients with mutations in ZFP57. *BMC Med. Genet.* **17**, 29 (2016).
- R. A. Waterland, R. Keller Mayer, E. Laritsky, P. Rayco-Solon, R. A. Harris, M. Travisano, W. Zhang, M. S. Torskaya, J. Zhang, L. Shen, M. J. Manary, A. M. Prentice, Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet.* **6**, e1001252 (2010).
- P. Dominguez-Salas, S. E. Moore, M. S. Baker, A. W. Bergen, S. E. Cox, R. A. Dyer, A. J. Fulford, Y. Guan, E. Laritsky, M. J. Silver, G. E. Swan, S. H. Zeisel, S. M. Innis, R. A. Waterland, A. M. Prentice, B. J. Hennig, Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat. Commun.* **5**, 3746 (2014).
- P. Dominguez-Salas, S. E. Moore, D. Cole, K.-A. da Costa, S. E. Cox, R. A. Dyer, A. J. C. Fulford, S. M. Innis, R. A. Waterland, S. H. Zeisel, A. M. Prentice, B. J. Hennig, DNA methylation potential: Dietary intake and blood concentrations of one-carbon metabolites and cofactors in rural African women. *Am. J. Clin. Nutr.* **97**, 1217–1227 (2013).
- C. Do, C. F. Lang, J. Lin, H. Darbary, I. Krupka, A. Gaba, L. Petukhova, J.-P. Vonsattel, M. P. Gallagher, R. S. Goland, R. A. Clynes, A. J. Dwork, J. G. Kral, C. Monk, A. M. Christiano, B. Tycko, Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *Am. J. Hum. Genet.* **98**, 934–955 (2016).

33. T. E. Van Baak, C. Coarfa, P.-A. Dugué, G. Fiorito, E. Laritsky, M. S. Baker, N. J. Kessler, J. Dong, J. D. Duryea, M. J. Silver, A. Saffari, A. M. Prentice, S. E. Moore, A. Ghantous, M. N. Routledge, Y. Y. Gong, Z. Herceg, P. Vineis, G. Severi, J. L. Hopper, M. C. Southey, G. G. Giles, R. L. Milne, R. A. Waterland, Epigenetic supersimilarity of monozygotic twin pairs. *Genome Biol.* **19**, 2 (2018).
34. R. A. Waterland, K. B. Michels, Epigenetic epidemiology of the developmental origins hypothesis. *Annu. Rev. Nutr.* **27**, 363–388 (2007).
35. A. P. Feinberg, R. A. Irizarry, Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1757–1764 (2010).
36. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
37. Y. Liu, K. D. Siegmund, P. W. Laird, B. P. Berman, Bis-SNP: Combined DNA methylation and SNP calling for bisulfite-seq data. *Genome Biol.* **13**, R61 (2012).
38. S. Guo, D. Diep, N. Plongthongkum, H.-L. Fung, K. Zhang, K. Zhang, Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **49**, 635–642 (2017).
39. A. E. Jaffe, P. Murakami, H. Lee, J. T. Leek, M. D. Fallin, A. P. Feinberg, R. A. Irizarry, Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int. J. Epidemiol.* **41**, 200–209 (2012).
40. S. E. Moore, A. J. C. Fulford, M. K. Darboe, M. L. Jobarteh, L. M. Jarjou, A. M. Prentice, A randomized trial to investigate the effects of pre-natal and infant nutritional supplementation on infant immune development in rural Gambia: The ENID trial: Early Nutrition and Immune Development. *BMC Pregnancy Childbirth* **12**, 107 (2012).
- provided by G. Hemani. We also thank the study participants from The Gambia and staff at Medical Research Council (MRC) Keneba, The Gambia. **Funding:** R.A.W. was supported by grants from the U.S. Department of Agriculture (CRIS 3092-5-001-059), the Cancer Prevention and Research Institute of Texas (RP170295), and the UK-MRC (MR/M01424X/1). The Gambian studies are supported by core funding MC-A760-5QX00 to the MRC Unit The Gambia/MRC International Nutrition Group by the UK MRC and the UK Department for the International Development (DFID) under the MRC/DFID Concordat agreement, by MRC grant MR/M01424X/1, and by Wellcome Trust grant WT086369MA. **Author contributions:** N.J.K. and M.J.S. analyzed the data. R.A.W. conceived the expanded ME screen and the idea of exploring developmental dynamics at MEs using the data of Guo *et al.* All the authors wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data sets supporting the conclusions of this article are available in GEO (bisulfite-seq reads from previous ME screen: GSM1091961, GSM1091962, GSM1091963, and GSM1091964; RRBS data from Guo *et al.*: GSE49828) and SRA (bisulfite-seq reads from the Roadmap Epigenomics Project: all data from experiments SRX388732, SRX175349, SRX263877, SRX388741, SRX388745, SRX109155, SRX263886, SRX263887, and SRX388747). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 8 February 2018

Accepted 1 June 2018

Published 11 July 2018

10.1126/sciadv.aat2624

Citation: N. J. Kessler, R. A. Waterland, A. M. Prentice, M. J. Silver, Establishment of environmentally sensitive DNA methylation states in the very early human embryo. *Sci. Adv.* **4**, eaat2624 (2018).

Acknowledgments: We thank S. E. Moore, B. J. Hennig, P. Dominguez-Salas, Z. Herceg, M. N. Routledge, Y. Y. Gong, and H. Hernandez-Vargas for providing samples and methylation data for the Gambian SoC analysis. Information on SNP heritability for the mQTL analysis was