Kessler, NJ; Waterland, RA; Prentice, AM; Silver, MJ (2018) Establishment of environmentally sensitive DNA methylation states in the very early human embryo. Science advances, 4 (7). eaat2624. ISSN 2375-2548 DOI: https://doi.org/10.1126/sciadv.aat2624

Downloaded from: http://researchonline.lshtm.ac.uk/4648534/

DOI: 10.1126/sciadv.aat2624

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
Establishment of environmentally sensitive DNA methylation states in the very early human embryo

Noah J. Kessler¹, Robert A. Waterland², Andrew M. Prentice¹, Matt J. Silver¹*

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.

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 (ERVK) elements [P = 4.7 × 10⁻⁸, P = 1.4 × 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
embryos using public data (20). This data set consists of reduced representative 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 < 2.2−16, 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. 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).
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 multifinger 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).
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 periconceptional 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 periconceptional nutritional environment (8, 11, 30). We therefore hypothesized that Gambian SoC-associated genomic regions...
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, periconceptional 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,
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 periconceptional 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, ENCCF002CRN; CTCF: H1–hESC (human embryonic stem cells), ENCCF002CD5]. 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 S3).

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 confidence 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 HumanMethylation450 BeadChip were downloaded from www.mqtldb.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

Table S5. Overlap of Bak

REFERENCES AND NOTES
1. H. Okae, H. Chiba, H. Hira, H. Hamada, A. Sato, T. Utsunomiya, H. Kikuchi, H. Yoshida, REFERENCES AND NOTES
2. M. D. Schultz, Y. He, J. W. Whitaker, M. Harhirkan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Liu, J. Jung, 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).
3. T. R. Gaunt, H. A. Shihab, G. Hemani, J. L. Min, G. Woodward, O. Lyttleton, J. Zheng, P. Liu, J. Yan, R. Ren, P. Yuan, Y. Yuan, Z. Yan, L. Wen, L. Yan, J. Qiao, Fang. Single-cell DNA methylation sequencing of human preimplantation embryos. Nature Genet. 50, 12–19 (2018).
4. A. L. He, T. Han, L. Chen, M.-L. Ong, S. K. Dogra, J. Wong, J. L. MacIsaac, S. M. Mah, M. L. McEwen, S.-M. Saw, K. M. Godfrey, Y.-S. Chong, K. Kwek, C.-K. Koh, S.-H. Koh, M. F. F. Chong, S. Barton, N. Kamani, C. Y. Cheong, J. P. Buschdorf, W. Stunkel, M. S. Kobor, M. J. Meany, P. D. Gluckman, J. D. Holbrook, The effect of genotype and uterine environment on interindividual variation in neonate DNA methylomes. Genome Res. 40, 1046–1074 (2014).
5. T. A. Rakyan, H. A. Shihab, G. Hemani, J. L. Min, G. Woodward, O. Lyttleton, J. Zheng, P. Liu, J. Yan, R. Ren, P. Yuan, Y. Yuan, Z. Yan, L. Wen, L. Yan, J. Qiao, Fang. Single-cell DNA methylation sequencing of human preimplantation embryos. Nature Genet. 50, 12–19 (2018).
6. A. C. Ferguson-Smith, Genomic imprinting: The emergence of an epigenetic paradigm. Nat. Rev. Genet. 12, 565–575 (2011).
7. R. A. Waterland, C. Garza, Potential mechanisms of metabolic imprinting that lead to chronic disease. Am. J. Clin. Nutr. 69, 179–197 (1999).
8. C. Faulk, J. H. Kim, O. S. Anderson, M. S. Nahar, T. R. Jones, M. A. Sartor, D. C. Dolinoya, A. C. Ferguson-Smith, Genomic imprinting: The emergence of an epigenetic paradigm. Nat. Rev. Genet. 12, 565–575 (2011).
9. M. D. Schultz, Y. He, J. W. Whitaker, M. Harhirkan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Liu, J. Jung, 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).
10. J. Ernst, M. Kellis, ChromHMM: Automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216 (2012).
11. C. Faulk, J. H. Kim, O. S. Anderson, M. S. Nahar, T. R. Jones, M. A. Sartor, D. C. Dolinoya, A. C. Ferguson-Smith, Genomic imprinting: The emergence of an epigenetic paradigm. Nat. Rev. Genet. 12, 565–575 (2011).
12. A. L. Hey, P. Han, L. Chen, M.-L. Ong, S. K. Dogra, J. Wong, J. L. MacIsaac, S. M. Mah, M. L. McEwen, S.-M. Saw, K. M. Godfrey, Y.-S. Chong, K. Kwek, C.-K. Koh, S.-H. Koh, M. F. F. Chong, S. Barton, N. Kamani, C. Y. Cheong, J. P. Buschdorf, W. Stunkel, M. S. Kobor, M. J. Meany, P. D. Gluckman, J. D. Holbrook, The effect of genotype and uterine environment on interindividual variation in neonate DNA methylomes. Genome Res. 40, 1046–1074 (2014).
33. T. E. Van Baak, C. Coarfa, P. -A. Duguid, 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).

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 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).