

ARTICLE

Received 7 Jan 2014 | Accepted 8 Sep 2014 | Published 10 Oct 2014

DOI: 10.1038/ncomms6187

Sexual dimorphism in epigenomic responses of stem cells to extreme fetal growth

Fabien Delahaye¹, N. Ari Wijetunga², Hye J. Heo¹, Jessica N. Tozour¹, Yong Mei Zhao¹, John M. Greally² & Francine H. Finstein¹

Extreme fetal growth is associated with increased susceptibility to a range of adult diseases through an unknown mechanism of cellular memory. We tested whether heritable epigenetic processes in long-lived CD34⁺ haematopoietic stem/progenitor cells showed evidence for re-programming associated with the extremes of fetal growth. Here we show that both fetal growth restriction and over-growth are associated with global shifts towards DNA hypermethylation, targeting *cis*-regulatory elements in proximity to genes involved in glucose homeostasis and stem cell function. We find a sexually dimorphic response; intrauterine growth restriction is associated with substantially greater epigenetic dysregulation in males, whereas large for gestational age growth predominantly affects females. The findings are consistent with extreme fetal growth interacting with variable fetal susceptibility to influence cellular ageing and metabolic characteristics through epigenetic mechanisms, potentially generating biomarkers that could identify infants at higher risk for chronic disease later in life.

¹ Department of Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Block Building, Room 631, Bronx, New York 10461, USA. ² Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Price Building, Room 322, Bronx, New York 10461, USA. Correspondence and requests for materials should be addressed to J.M.G. (email: john.greally@einstein.yu.edu) or to F.H.E. (email: francine.einstein@einstein.yu.edu).

nvironmental factors have the potential for significant impact on normal development and health throughout the life span. Suboptimal intrauterine conditions represent a specific type of environmental exposure that is associated with increased risk for cardiovascular disease^{1,2} and premature death in adulthood³. A substantial amount of evidence has also demonstrated the relationship between poor maternal nutrition or low birth weight with a range of metabolic disorders and obesity in humans^{4,5} with animal studies further corroborating these findings⁶. At the opposite end of the spectrum of extreme fetal growth, excess nutrition leading to large for gestational age (LGA) birth weights is associated with similar adult phenotypes, with increased risk for premature mortality³ and a range of other age-related diseases⁷. Fetal growth restriction and over-growth show a decline in resistance to chronic disease in adulthood and involvement of multiple organ systems, which is typical of normal ageing and may represent a precocious ageing phenotype associated with both extremes of the fetal growth spectrum⁸.

Adverse exposures appear to be particularly consequential in early life, possibly due to the rapid expansion of cell populations necessary for growth, and the dynamism of cellular differentiation and lineage commitment that occurs during this period of development. Inherent to the differentiation process is the modification of transcriptional regulatory patterns. These include epigenetic regulators that are capable of transmitting newly established regulatory marks through cell replication⁹. Environmentally induced perturbations of the cell's normal epigenetic regulatory controls may be maintained in long-lived, self-renewing cells, maintained through proliferation and resulting in functional consequences later in life. Although alterations in DNA methylation has been associated with the cumulative exposures inherent to ageing 10, environmental exposures early in life may induce addition dysregulation of the epigenome conferring increased susceptibility for age-related

disease at a younger age.

We^{11,12} and others^{13–15} have explored the possibility that nonrandom epigenetic changes are associated with intrauterine growth restriction (IUGR). In studies of disease or phenotypeassociated epigenetic changes, the choice of cell type generally represents a compromise between accessibility, purity, quantity and mechanistic relevance. Unpurified peripheral blood leukocytes have previously been studied in individuals whose mothers were exposed prenatally to famine. Altered DNA methylation of multiple sites within the differentially methylated region of the imprinted Insulin-like growth factor 2 (IGF2) gene was found in subjects decades later 13. Cord blood leukocytes have also been used to demonstrate associations of DNA methylation with in utero conditions and birth weight^{15,16}. Another commonly studied tissue type is the placenta, which

functions at the maternal-fetal interface and may be a potential of intrauterine environmental conditions^{17–19} However, testing the placenta does not address the latent risk in adulthood of chronic disease, which has to be mediated by somatic cells of the offspring. Furthermore, the use of samples of mixed populations of cells in DNA methylation studies, such as those sampled from highly heterogeneous placental tissue, is now recognized as a major source of experimental artefact that limits interpretability of results^{20,21}.

We focus on haematopoietic stem/progenitor cells (HSPCs), purified using the CD34 surface marker to reduce cell subtype effects²⁰. HSPCs include a subset of long-term, self-renewing stem cells that persist through the life of the individual²², allowing the cellular propagation or the 'memory' of exposure to temporally remote suboptimal conditions. The role of CD34⁺ HSPCs in the maintenance of vascular integrity^{23,24} is mechanistically relevant for the adult phenotype associated with increased risk for cardiovascular disease^{4,25}. We studied infants born with the two extremes of fetal growth, IUGR and LGA, compared with control infants with appropriate weight for gestational age. Owing to the thorough characterization of CD34⁺ HSPCs by the Roadmap Epigenomics Program, we were able to exploit the mapping of chromatin constituents to define empirically the cis-regulatory elements, such as promotors and enhancers, specific to this cell type²⁶ allowing us to interpret changes in DNA methylation at otherwise unannotated loci in the genome.

Results

Genome-wide DNA methylation profiling. We perform genome-wide DNA methylation profiling on purified CD34⁺ HSPCs from 60 subjects, 20 in each of three groups defined by appropriate or excessively large or small birth weight and ponderal index for gestational age and sex (Table 1). The HELP-tagging assay is used as a survey technique testing ~1.8 million loci quantitatively at nucleotide resolution and including relatively CG dinucleotidedepleted loci²⁷. This assay generates a methylation score that is inversely correlated to DNA methylation level, with a methylation score of 0 indicating full methylation and 100 indicating complete lack of methylation, based on a normalized ratio between tag counts generated by the methylation-sensitive enzyme HpaII and its methylation-insensitive isoschizomer MspI²⁸. Based on quality control measures (Methods and Supplementary Fig. 1), 993,514 loci are selected for further analyses. Of these, 10,043 loci are defined as candidate differentially methylated loci using batch-adjusted significance and degree of methylation difference thresholds in comparisons of IUGR and LGA infants with the normal birth weight controls. We observe a global relative shift towards DNA hypermethylation in CD34+ HSPCs in both

Table 1 Clinical cohort characteristics.						
	Cohort 1			Cohort 2		
	Control (n = 20)	IUGR (n = 20)	LGA (n = 20)	Control (n = 8)	IUGR (n = 8)	LGA (n=8)
Gestational age, weeks	39.3 ± 0.3	38.7 ± 0.4	39.2 ± 0.2	39.1 ± 0.5	38.7 ± 0.4	39.8 ± 0.3
Birth weight, g	$3,150 \pm 64$	2,515 ± 69*	3,996 ± 81*	$3,150 \pm 161$	$2,498 \pm 78^{\dagger}$	4,053 ± 67*
Ponderal index, g cm ⁻³	2.8 ± 0.03	2.3 ± 0.04 *	$3.2 \pm 0.03^{*}$	2.8 ± 0.07	2.3 ± 0.04 *	$3.4 \pm 0.07^*$
% Male	50	50	50	50	50	50
Maternal age, years	24.6 ± 0.9	26.5 ± 1.3	31.1 ± 1.1*	33.1 ± 1.7	$27.7 \pm 2.6^{\dagger}$	31.5 ± 1.6
Pre-pregnancy BMI, kg m ⁻²	26.6 ± 1.3	25.5 ± 0.9	29.0 ± 1.3	27.0 ± 2.1	26.5 ± 2.8	28.5 ± 1.4

29.9 + 2.1

 23.0 ± 2.7

BMI, body mass index; IUGR, intrauterine growth restriction; LGA, large for gestational age

29.4 + 2.7

Weight gain, pounds

 27.0 ± 5.1

12.7 + 5.1

24.8 + 2.9

Showing mean + standard deviation values

^{*}P<0.001 compared with Control.

IUGR and LGA subjects when compared with the controls (Fig. 1a). The clustering of cases (LGA/IUGR) is not uniform, with a subset of cases clustering with controls (Fig. 1b), indicating that epigenetic dysregulation does not occur universally as a response to extreme fetal growth. Although there exists a subset of common loci altered in both IUGR and LGA neonates, most of the dysregulated loci are distinctive between these groups (Fig. 1c,d). We also see an overlap of genes (as opposed to loci) undergoing differential DNA methylation (Supplementary Fig. 2).

Sexual dimorphism associated with the extremes of fetal growth. Sex-specific comparisons for DNA methylation patterns are shown between control and IUGR and LGA subjects (Fig. 2). Both IUGR males and females show a shift in DNA methylation profiles compared with controls, but the number of hypermethylated loci is markedly higher in males compared with females (Fig. 2a). Sex-specific differences are also seen in the comparison of LGA to controls, with LGA females showing an increase in the overall number of candidate differentially methylated loci compared with males (Fig. 2b). These findings indicate a sexual dimorphism in the epigenetic responses of HSPCs to the extremes of growth conditions *in utero*.

Targeting of DNA methylation changes to specific genomic contexts. Although the consequences of DNA methylation changes at recognized promoter sequences are generally predictable, a genome-wide study of this type can generate a majority of findings in un-annotated genomic locations. To predict the

functional consequences of these candidate differentially methylated loci, we take advantage of the mapping of chromatin components in CD34⁺ HSPCs performed as part of the Roadmap Epigenomics Program. The details of this annotation are described in a separate report²⁶ and involve the use of the Segway algorithm²⁹ to generate genomic features (Methods) that are then interpreted using Self-Organizing Maps³⁰. We are thus able to define candidate promoters, enhancers, transcribed sequences and repressive chromatin in the epigenome specific to the CD34 + HSPC population. Every HpaII site is then assigned to a candidate feature based on its genomic position. The HELPtagging assay represents each of the candidate genomic features (based on 993,514 loci) and the candidate differentially methylated loci (10,043) are significantly enriched in Segway features 4 (enhancers, P < 0.001) and 6 (promoters, P < 0.001), indicating preferential targeting to transcriptional regulatory elements (Fig. 3a). We show an example of the mapping of the one the candidate differentially methylated loci, to the promoter of the Retinoid X receptor, alpha (RXRA) gene, at an annotated CpG island, and within the Segway feature 6 annotation indicating candidate promoter function. The HELP-tagging derived methylation scores for cases (IUGR and LGA combined) are compared with controls to demonstrate the magnitude of the change at this locus (Fig. 3b).

Targeting of DNA methylation changes to genes with specific properties. We test whether the subset of loci affected by DNA methylation changes are enriched at a specific subset of genes characterized by concordance of function of their protein

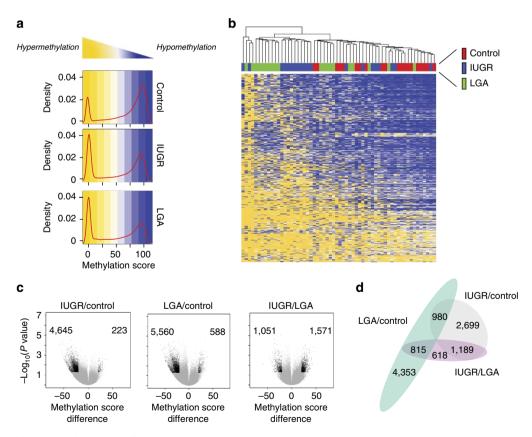


Figure 1 | Genome-wide DNA methylation profiles. (a) Density plots of methylation scores for IUGR or LGA compared with controls. The distributions of DNA methylation scores are shown in red. (b) A self-organizing heatmap of candidate differentially methylated loci showing clustering by sample. (c) Volcano plots of DNA methylation score differences for IUGR compared with control, LGA compared with control and IUGR compared with LGA, based on 993,514 loci throughout the genome. Differentially methylated loci with *P* value < 0.05 and methylation difference > |20| are shown in black. (d) Differentially methylated loci meeting threshold criteria are quantified in a proportional Venn diagram for each comparison.

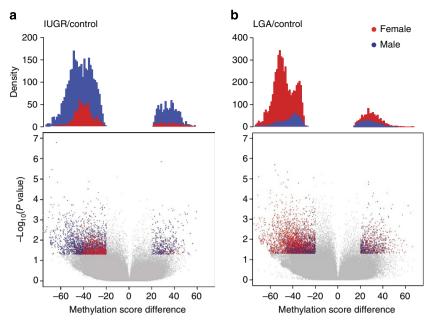


Figure 2 | Sexual dimorphism in IUGR males and LGA females for differentially methylated loci. The lower panels show volcano plots of DNA methylation score differences, the upper panels quantify the densities of differentially methylated loci (*P* value < 0.05 using analysis of variance with pairwise two-tailed Tukey-tests, methylation difference > |20|). (a) IUGR compared with controls, (b) LGA compared with controls.

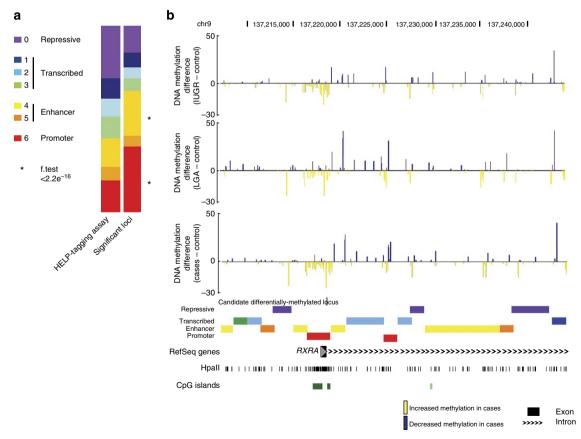


Figure 3 | Candidate differentially methylated loci are enriched at *cis***-regulatory elements.** (a) Based on empirical annotation of promoter, enhancer, repressive and transcribed regions, enrichment of candidate differentially methylated loci (n = 10,043) in cases (IUGR and LGA) compared with controls is illustrated with significance values shown for enriched sequence features. The bar on the left represents the proportional representation of each feature in terms of loci tested by HELP-tagging, whereas the bar on the right shows the proportions of features at which differentially methylated loci are found. Significant enrichment for differential methylation at candidate promoters and enhancers is observed. (b) An example of the *RXRA* gene with a candidate differentially methylated locus is shown. The DNA methylation score differences between controls and IUGR (top), LGA (middle) and cases (bottom, IUGR and LGA combined) are depicted, with a site identified as being a candidate differentially methylated locus in the CpG island promoter region shown in grey. Blue, positive values represent decreased DNA methylation in the cases of extreme fetal growth; yellow, negative value increased methylation.

products. A candidate differentially methylated locus is linked to a specific gene if the site is (i) located in proximity to the transcription start site of the RefSeq gene and (ii) overlapping candidate regulatory loci (features 4 or 6). We select only those candidate promoters (feature 6) within ±2kb and candidate enhancers (feature 4) within ± 5 kb of RefSeq transcription start sites. Although enhancers can act over substantially longer distances than 5 kb (ref. 31), we are deliberately conservative in restricting the distance so that we would be more likely to associate an enhancer with the gene upon which it exerts its effects. The resulting list of genes is used to perform a gene set enrichment analysis (GSEA). Traditional GSEA does not take into account the physical characteristics of the gene and has been shown to be biased by factors such as the numbers of CG dinucleotide sites associated with different classes of gene and gene promoters³². To address this, the Bioconductor package GoSeq³³ was developed to control for variability of length of genes. We adapted GoSea to normalize our data to control for the number of CG dinucleotides linked to each gene by the above criteria. Detailed information describing the results of the normalized GSEA is shown in Supplementary Tables 1 and 2. Among the different significant pathways from KEGG (Kyoto Encyclopedia of Genes and Genomes), two pathways of interest emerge as significant regardless of group comparison: the KEGG pathways for Maturity onset diabetes of the young, relevant to glucose homeostasis and Hedgehog (HH) signalling. Both of these pathways contain genes involved in proliferation, differentiation and self-renewal capabilities of stem cells. Permutation analysis was performed to confirm the significance of these results. Based on the criteria for assigning HpaII sites to RefSeq genes described above, the HELP-tagging assay represents 97.6% of RefSeq genes, so we randomly select from within this group of genes the same number of genes used to define our pathways, and perform the GSEA analysis 1,000 and 3,000 times to test how frequently the same pathways are identified as, defining the significance of our detection of these pathways as $P < 10^{-3}$. The same pathways are targeted by IUGR and LGA even when the loci involved are not identical (Fig. 4). A similar effect is seen for the loci affected differentially between males and females (Supplementary Fig. 3). These findings combine to show convergence of dysregulation of the same pathways by IUGR and LGA and in male and female subjects,

respectively, even though the loci targeted for DNA methylation changes are not necessarily the same in each group.

Verification and validation. To test the robustness of our genome-wide technique, we assess the reproducibility of DNA methylation differences at our candidate differentially methylated loci using single-locus quantitative validation studies. We first perform verification studies on samples from Cohort 1, on whom the genome-wide studies had been performed, testing four loci selected for differing levels of DNA methylation in 24 subjects. A strong correlation between bisulphite MassArray and HELP-tagging is found ($R^2 = 0.98$, Supplementary Fig. 4). In a second, independent set of CD34⁺ HSPC samples (Cohort 2) consisting of eight new subjects per group (control, IUGR, LGA) with equal numbers of males and females in each group, we perform a targeted bisulphite sequencing (TBS) assay, using bisulphite treatment, targeted PCR and massively parallel sequencing to measure DNA methylation at 72 loci in the 24 subjects (see Methods and Supplementary Table 3). The correlations between HELP-tagging with MassArray in Cohort 1 ($R^2 = 0.98$, Supplementary Fig. 4) and with TBS in Cohort 2 ($R^2 = 0.72$, Supplementary Fig. 5 and Supplementary Table 3) are both strong. These highly quantitative verification and validation studies demonstrate the technical robustness of the genome-wide HELP-tagging assay, as well as the potential to validate DNA methylation differences, even when using a new cohort of subjects. Of the 54 candidate differentially methylated loci from the HELP-tagging group comparisons, we focus on loci implicated by our GOseq-normalized GSEA results, using primers for candidate differentially methylated loci proximal to WNT6 (Fig. 5a) and PTCH1 (Fig. 5b) from the HH signalling pathway and MAFA from the Maturity onset diabetes of the young pathway (Supplementary Fig. 6). We find the direction of DNA methylation changes to be concordant between genome-wide and targeted assays for all three loci, with statistically significant differences demonstrable for TBS data from WNT6 (P = 0.023) and PTCH1 (P = 0.014; Supplementary Table 4). We also show the PTCH1 and WNT6 genes to have increased DNA methylation by TBS at local cis-regulatory elements in cases (IUGR and LGA) compared with controls (Supplementary Table 4, P < 0.05). Finally, we interrogate loci associated with genes that are differentially methylated on average between cases (IUGR plus LGA) and

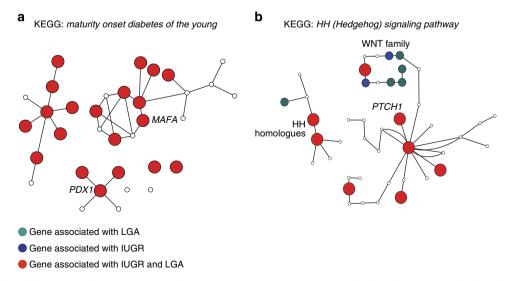


Figure 4 | Network analysis. A network representation of KEGG pathways for (a) *Maturity onset diabetes of the young* and (b) *Hedgehog (HH) signalling*. Nodes are colour- and size-coded based on the association of genes represented by each node with LGA or IUGR, or with both LGA and IUGR. Edges (solid lines) represent known physical interaction between genes.

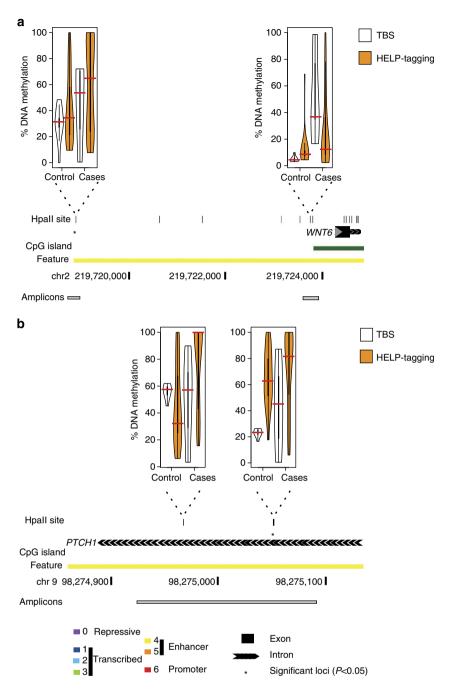


Figure 5 | Biological validation. Validation of significant loci of interest by targeted bisulphite sequencing (TBS) in Cohort 2 for loci at the (a) *WNT6* and (b) *PTCH1* genes. Candidate differentially methylated loci are shown as the Hpall sites within the amplicon regions (grey boxes), with results of DNA methylation distributions for controls and cases (IUGR and LGA combined) from HELP-tagging (orange) and TBS (white) depicted as violin plots (mean shown in red, 1st and 3rd quartile are depicted by the thick black bar). The results show concordance for similar types of changes between HELP-tagging and TBS results at these loci with significant results (*P*<0.05, *t*-test) for TBS marked with asterisks.

controls in Cohort 1 and previously found to have epigenetic alterations related to metabolic syndrome and type 2 diabetes mellitus, *IGF2* (refs 34,35) and *RXRA*³⁶. A positive correlation is seen between the HELP-tagging and TBS DNA methylation levels, but the TBS DNA methylation differences between cases and controls are of insufficient magnitude for statistical significance to be attributed (Supplementary Table 4). The number of new samples in Cohort 2 is limited and thus these validation studies are likely to confirm loci of major effect only. Overall, the TBS data are concordant with the genome-wide data, indicating that the conclusions based on the genome-wide results are tenable.

Discussion

Here we show for the first time epigenetic changes associated with the two very different types of intrauterine conditions reflected by extremes of fetal growth. All subjects were healthy, full-term neonates without any anomalies or dysmorphic features that would suggest an etiology innate to the fetus. We combined birth weight with ponderal index values to define the distinct groups of infants at either end of the fetal growth continuum. Both groups with abnormal growth demonstrate global shifts of increased DNA methylation compared with appropriately grown neonates. Although the underlying differences in environmental exposures

cannot be measured with precision in these subjects, the intrauterine conditions of those born at the extremes of birth weight are likely to differ substantially from each other. Despite this, the DNA methylation profiles of infants at both ends of the growth spectrum are more similar to each other than to the control subjects. Several factors contribute to the strength of this study. First, the two-stage design of this study increases confidence in our findings. In a recent review of 257 epigenome-wide association studies (EWASs) the median number of study subjects included was 46, with only about one-third of studies validating results in a second cohort²¹. Our technical verification and then validation in Cohort 2 illustrate the robustness of the predictions from the genome-wide assays, increasing confidence in our results. We tested CD34⁺ HSPCs as cells with both the long-lived properties and mechanistic properties in inflammation and maintenance of vascular integrity that make it a potential mediator of adult disease risks associated with extreme fetal growth.

We find that IUGR and LGA share a common response with a tendency towards increased DNA methylation. The targeted loci show enrichment at candidate *cis*-regulatory elements and proximity to genes encoding proteins with functions implicated in the *Maturity onset diabetes of the young* and *HH* signalling pathways, in both IUGR and LGA subjects despite sharing only a subset of identical loci undergoing DNA methylation dysregulation. These gene/protein properties are significant when considered in terms of the adult phenotype associated with abnormal fetal growth, such as premature glucose intolerance and type 2 diabetes mellitus³⁷. HH signalling is critical for stem cell proliferation and self-renewal³⁸, is necessary for haematopoietic stem cell fate decisions³⁹ and may play a critical role in CD34 + cells reparative contributions after myocardial infarction⁴⁰.

We also find a sexual dimorphism in the DNA methylation profiles, with IUGR males and LGA females showing greatest alterations in global DNA methylation. Some of the large epidemiological studies that examined outcomes for males and females separately have found sex-specific differences^{2,41,42}, although these have not been consistently reported. Results from other EWAS studying epigenetic responses to adverse intrauterine conditions have generally presented combined observations for males and females 18,19. Our findings of a global shift towards increased DNA methylation is inconsistent with findings seen when the relationship between fetal growth and DNA methylation of long interspersed nuclear elements (LINE-1) was tested in cord blood leukocytes^{15,43} and placenta¹⁴. Decreased LINE-1 methylation, which has been associated with genomic instability and cancer risk44, was found in cord blood from newborns with low and high birth weight⁴³. The development of adiposity in 5- to 12-year-old boys but not girls has also been associated with decreased LINE-1 methylation in peripheral leukocytes⁴⁵. Others have examined sex-specific changes in association with other environmental exposures, but these were either in a limited number of differentially methylated regions⁴⁶ or global changes with limited sample size numbers^{47,48}. All of the prior studies included mixed cell-type samples, which can hamper interpretation of DNA methylation studies²⁰.

To avoid the possibility of artifactual results stemming from testing of mixed cell populations, we examined purified cell samples. CD34⁺ HSPCs were chosen for their self-renewal properties that enable them to propagate a cellular memory of temporally remote events, and for their mechanistically plausible contribution to the associated adult phenotype, especially the increased susceptibility to cardiovascular complications. CD34⁺ HSPCs play a key role in maintaining the intravascular endothelial layer. Intimal denudation generally precedes the development of atherosclerosis⁴⁹ and HSPCs contribute to repair

after peripheral ischemic injury through differentiation into endothelial cells⁵⁰ but may also mediate repair through stem cell paracrine effects^{51,52}. In adults, circulating numbers of CD34⁺ HSPCs have been shown to be inversely related to cardiovascular disease risk^{23,24}. Furthermore, HSPCs can be induced to differentiate into multiple tissue types, including those involved in metabolic regulation^{53,54}. Impaired mobilization of HSPCs from the bone marrow⁵⁵ and decreased circulating HSPCs⁵⁶ are thought to link metabolic disorders such as diabetes to cardiovascular disease risk. The effect of ageing on HSPCs includes reduced self-renewal capacity^{57,58} and increased myeloid-biased differentiation^{57,58}, which has been associated with the increased susceptibility to chronic age-related diseases ^{59,60}. HSPCs from young and old individuals are similarly effective in reconstituting blood lineages after transplantation^{58,61}, but aged HSPCs may be less effective at homing and engrafting at the sites of injury⁵⁷. The regenerative potential of haematopoietic stem cells is multifaceted and the contributing roles of functional defects in the stem cell population itself versus impairment of the tissue environment are as yet unknown. Future studies are needed to determine the clinical impact for stem cell transplantation when using cord blood samples from these otherwise healthy neonates with abnormal fetal growth, as our study suggest that HSPCs may decrease their abilities to renew and differentiate after exposure.

Studying the epigenetic basis for developmental origins of adult disease poses several challenges because of the inaccessibility of the human fetus in utero, the lack of tools to measure intrauterine exposures over course of the pregnancy and the duration of the time needed to study outcomes that evolve over decades of the human lifespan. Although no study is without limitations, we present our work in part as a framework for discussion of the challenges and considerations when designing EWAS in the future. Our findings indicate that sex-specific differences should be examined in addition to a range of clinical phenotypes or experimental intrauterine exposures in animal models. CD34+ HSPCs represent a homogenous, accessible cell population directly relevant to the study of developmental origins of adult disease given their known involvement in cardiovascular disease risk. CD34⁺ HSPCs are also well characterized from a genomic perspective through Roadmap Epigenomics Program mapping studies, allowing the investigator to interpret findings in terms of functional elements in this cell type specifically. Defining the functional implications of perturbations in the pathways identified here will be a valuable further avenue of research. Our findings provide key insights into how seemingly opposing intrauterine exposures give rise to similar adult phenotypes, through perturbation of DNA methylation converging on common loci or at distinct loci targeting genes in common pathways. As methods for design and execution of EWAS become better defined, the discovery of novel biomarkers that represent cumulative prior exposures in early life may ultimately provide new tools that identify at-risk neonates for preventative interventions.

Methods

This study was approved by the Institutional Review Board of the Montefiore Medical Center and the Committee on Clinical Investigation at the Albert Einstein College of Medicine and is in accordance with Health Insurance Portability and Accountability Act regulations. Written informed consent was obtained from all subjects before participation.

Sample collection. Cord blood from neonates was the source of material for this study. Biological samples and clinical information were collected (n=84) from consenting women who delivered healthy infants without any anomalies or dysmorphic features and following an uncomplicated intrapartum course, without evidence of fetal distress (normal Apgar scores and cord blood gases without acidemia). The three groups were comprised of infants with appropriate growth,

IUGR or LGA (matched for gestational age at delivery and sex). Both birth weight and ponderal index (a measurement of neonatal weight relative to length) were used to identify case and control subjects. IUGR and LGA were, respectively, defined by birth weight and ponderal index values <10th or >90th percentile for gestational age and sex. Control infants had normal parameters (>10th and <90th percentiles) for both birth weight and ponderal index. Maternal and infant characteristics are shown in Table 1. Cohort 1 (genome-wide assays) is composed of 20 samples per group, whereas Cohort 2 (validation cohort, targeted assays) has 8 samples per group, with equal representation between male and female subjects in all groups.

Isolation of CD34 $^+$ **HSPCs.** CD34 $^+$ cells, which constitute approximately 1% of nucleated blood cells in umbilical cord blood⁶², were isolated from the cord blood specimen using an immunomagnetic separation technique. Mononuclear cells were separated by Ficoll-Paque density gradient or using PrepaCyte-WBC following which CD34 $^+$ cells were obtained by positive immunomagnetic bead selection, using the AutoMACS Separator (Miltenyi Biotech). This resulted in the isolation of cells with \geq 95% purity. We cryopreserved the purified cells in 10% dimethyl sulphoxide using controlled rate freezing.

Genome-wide DNA methylation assay. The HELP-tagging assay was performed after isolation of genomic DNA from frozen CD34 $^+$ HSPCs, digested to completion by either HpaII or MspI. The digested DNA was ligated to two custom adapters containing Illumina adapter sequences, an EcoP15I recognition site and the T7 promoter sequence. Using EcoP15I, we isolated sequence tags flanking the sites digested by each enzyme, methylation-sensitive HpaII or methylation-insensitive MspI, followed by massively parallel sequencing of the resulting libraries (Illumina Technology)²⁷. HpaII profiles were obtained for each sample (n=60), calculating methylation scores using a previously generated MspI human reference.

Data processing and statistical analysis. DNA methylation scores from 0 (fully methylated) to 100 (unmethylated) were filtered by confidence scores. These confidence scores were calculated for each sample based on the total number of HpaII-generated reads as a function of the total number of MspI-generated reads, excluding loci for which the confidence score was lower than the expected mean by locus. To understand the relative effects of known technical covariates acting on methylation data variability, we performed principal components analysis (R package *princomp*) on the DNA methylation score obtained from the preprocessed data. We fit a linear model for each of the ten principal components as a function of each covariate, and summarized the data with a heatmap of the negative $\log_{10} P$ -values of each regression. We found batch effects (date of sequencing and presence in the same lane of the Illumina machine) to be significant confounding covariates (Supplementary Fig. 7). We confirmed that the effect of a global increase of DNA methylation in cases compared with controls remained after controlling by adjusting P values for the batch covariate.

Candidate differentially methylated loci were identified using analysis of variance with pairwise two-tailed Tukey-tests when comparing controls with either IUGR or LGA as well as two-sided t-tests when comparing control/cases to define locus-specific differences in average methylation between groups. Confirmatory linear regression of DNA methylation on group adjusting for batch effects was also performed. Comparisons between control/IUGR, control/LGA and IUGR/LGA were also stratified by sex. For each comparison, only loci with at least eight samples from each sex were retained. After confidence score filtering and selection for a minimum number of observations in each group, the number of testable loci decreased from >1.8 million to 993,514. We defined candidate differentially methylated loci to have a difference between mean DNA methylation scores > 20% and a $\stackrel{'}{P}$ value < 0.05. The necessary amplitude of DNA methylation score differences was defined using power calculations from a preliminary analysis of a subset of our samples (five per group). Using the average methylation from the control group, the standard deviations from each of the three groups and 16 samples per group (our minimum sample requirement per group after quality control measures), we are fully powered (>99%) to detect at least one group methylation difference > 20% at an false discovery rate (FDR) = 0.05. To demonstrate that our technique still exceeds minimum power recommendations for gender-specific comparisons, we ran simulations of the gender comparisons with eight samples per group and we were powered at roughly 85% to detect at least one group methylation difference > 20% at an FDR = 0.05.

Bisulphite MassArray verification assays. We selected 24 samples from Cohort 1 to test the technical performance of the genome-wide DNA methylation studies, a verification approach on our original cohort. Bisulphite conversion and MassArray (Sequenom) were performed⁶³. Primers were designed to cover loci with low, intermediate and high levels of DNA methylation from the HELP-tagging data across all samples regardless of group (Supplementary Fig. 4).

Targeted bisulphite sequencing. We bisulphite-converted 200 ng of DNA using the Zymo EZ-96 Methylation-Lightning Kit. After separate PCR amplification of individual target regions (primers listed in Supplementary Table 5), we pooled the

amplicons in equal ratios and generated Illumina libraries using robotic automation (Tecan). In total, 24 libraries were multiplexed on the Illumina Miseq for 250 bp paired end sequencing. Amplicons were selected to be part of our differentially methylated loci, covering the entire spectrum of predicted DNA methylation values (from 0 to 100) and allowing us to validate our specific pathways.

Amplicon bisulphite sequence alignment, DNA methylation calls. Sequence reads from the Illumina MiSeq were trimmed for adapter sequences and aligned to the human genome using bsmap (Bisulphite Sequencing Mapping Platform)⁶⁴ using the default settings, requiring a PHRED score of \geq 37 during alignment. We checked for bisulphite conversion efficiency (C \rightarrow T in CH contexts, Supplementary Table 6) and quantified the percent methylation for each sample (from 0 (unmethylated) to 1 (fully methylated)) at every CpG in the amplicons using the methratio tool provided by bsmap. We performed validation on 24 new CD34 $^+$ HSPC samples (Cohort 2 with eight subjects per group).

Genome annotation. We obtained publicly available chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) data from the Roadmap in Epigenomics project for CD34 + mobilized HSPCs from a 33-year-old, Caucasian female (RO_01549/GSM706857). Annotation of genomic features consisted of processing raw data provided through http://www.roadmapepigenomics.org/ for chromatin accessibility (DNase hypersensitivity) as well as ChIP-seq data for six histone modifications, followed by the use of the Segway algorithm²⁹ to predict seven features, interpreted using Self-Organizing Maps³⁰ and RefSeq gene metaplots to define promoter, enhancer, transcribed and repressed sequences in the CD34 + HSPCs.

Functional enrichment analysis. To perform GSEA⁶⁵ we first linked RefSeq genes to our candidate differentially methylated loci. We filtered these candidate differentially methylated loci to include only those overlapping candidate promoters (feature 6) or enhancers (feature 4), thereby enriching for loci with greater likelihood to have functional consequences. Candidate differentially methylated loci overlapping candidate promoters within ±2 kb and candidate enhancers within ± 5 kb of RefSeq gene transcription start sites were used to link DNA methylation changes with specific genes. Differentially enriched pathways found using a FDR q value < 0.05 are shown in Supplementary Table 1. We found using a FDR q value < 0.05 are snown in suppression, 1 validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the KEGG Maturity onset diabetes of the young and HH signalling validated the young and young and HH signalling validated the young and young and young and HH signalling validated the young and you pathways (Supplementary Table 2) using the Bioconductor package GOseq control for bias because of the variation of number of HpaII sites associated with different genes. As our analyses included a large number of genes, we wanted to test further the robustness of the enrichment of the two pathways selected, by generating random data sets using a permutation approach. As our original analysis was based on the top 2,000 candidate differentially methylated loci from HELP-tagging assay, we selected 2,000 genes randomly from those represented by HELP-tagging (97.6% of total) from the hg19 RefGene database (R package geneLenDataBase, database org.Hs.egREFSEQ2EG) 1,000 or 3,000 times, and ran the GOseq algorithm on each of these samples. The Maturity onset diabetes of the young pathway was not predicted in any of the 1,000 iterations, and three times in the 3,000 iterations, whereas HH signalling was predicted once in the 1,000 iterations and not at all in the 3,000 iterations. We therefore define the observed enrichment for these pathways at our dysregulated genes to be specific and statistically significant (P < 0.001). We visualize the association of DNA methylation changes and gene properties using pathways identified from the KEGG (http://www.genome.jp/kegg/pathway.html). Gene pathways were visualized in Cytoscape v3.0.2 with edges representing the physical interactions between nodes (genes/proteins). Node colours and sizes were adjusted to reflect the enrichment for IUGR or LGA separately or together (Fig. 4), and for sex specificity (Supplementary Fig. 3). A complete list of genes associated with these pathways is shown in Supplementary Table 7.

References

- Barker, D. J. & Osmond, C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet* 1, 1077–1081 (1986).
- Roseboom, T. J. et al. Coronary heart disease after prenatal exposure to the Dutch famine, 1944-45. Heart 84, 595-598 (2000).
- Baker, J. L., Olsen, L. W. & Sorensen, T. I. Weight at birth and all-cause mortality in adulthood. *Epidemiology* 19, 197–203 (2008).
- Barker, D. et al. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidemia (syndrome X): relation to reduced fetal growth. *Diabetologia* 36, 62–67 (1993).
- Olsen, L. W., Baker, J. L., Holst, C. & Sorensen, T. I. Birth cohort effect on the obesity epidemic in Denmark. *Epidemiology* 17, 292–295 (2006).
- Seki, Y., Williams, L., Vuguin, P. M. & Charron, M. J. Minireview: Epigenetic programming of diabetes and obesity: animal models. *Endocrinology* 153, 1031–1038 (2012).
- Dyer, J. S. & Rosenfeld, C. R. Metabolic imprinting by prenatal, perinatal, and postnatal overnutrition: a review. Semin. Reprod. Med. 29, 266–276 (2011).

- Thompson, R. F. & Einstein, F. H. Epigenetic basis for fetal origins of agerelated disease. J. Womens Health (Larchmt) 19, 581–587 (2010).
- Cedar, H. & Bergman, Y. Programming of DNA methylation patterns. Annu. Rev. Biochem. 81, 97–117 (2012).
- Bocker, M. T. et al. Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. Blood 117, e182–e189 (2011).
- 11. Einstein, F. et al. Cytosine methylation dysregulation in neonates following intrauterine growth restriction. PLoS One 5, e8887 (2010).
- Thompson, R. F. et al. Experimental intrauterine growth restriction induces alterations in DNA methylation and gene expression in pancreatic islets of rats. J. Biol. Chem. 285, 15111–15118 (2010).
- Heijmans, B. T. et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc. Natl Acad. Sci. USA 105, 17046–17049 (2008).
- 14. Bourque, D. K., Avila, L., Penaherrera, M., von Dadelszen, P. & Robinson, W. P. Decreased placental methylation at the H19/IGF2 imprinting control region is associated with normotensive intrauterine growth restriction but not preeclampsia. *Placenta* 31, 197–202 (2010).
- Fryer, A. A. et al. Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. Epigenetics 6, 86–94 (2011).
- Kile, M. L. et al. Correlation of global and gene-specific DNA methylation in maternal-infant pairs. PLoS One 5, e13730 (2010).
- Bouchard, L. et al. Placental adiponectin gene DNA methylation levels are associated with mothers' blood glucose concentration. Diabetes 61, 1272–1280 (2012).
- Ruchat, S. M. et al. Gestational diabetes mellitus epigenetically affects genes predominantly involved in metabolic diseases. Epigenetics 8, 935–943 (2013).
- Nomura, Y. et al. Global Methylation in the placenta and umbilical cord blood from pregnancies with maternal gestational diabetes, preeclampsia, and obesity. Reprod. Sci. 21, 131–137 (2013).
- Houseman, E. A. et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13, 86 (2012).
- Michels, K. B. et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat. Methods 10, 949–955 (2013).
- Dykstra, B. et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell 1, 218–229 (2007).
- Taguchi, A. et al. Circulating CD34-positive cells provide an index of cerebrovascular function. Circulation 109, 2972–2975 (2004).
- 24. Valgimigli, M. et al. CD34⁺ and endothelial progenitor cells in patients with various degrees of congestive heart failure. Circulation 110, 1209–1212 (2004).
- Barker, D. J., Osmond, C., Golding, J., Kuh, D. & Wadsworth, M. E. Growth in utero, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *BMJ* 298, 564–567 (1989).
- Wijetunga, N. A. et al. The meta-epigenomic structure of purified human stem cell populations is defined at cis-regulatory sequences. Nat. Commun. 5, 5195 (2014).
- Suzuki, M. et al. Optimized design and data analysis of tag-based cytosine methylation assays. Genome Biol. 11, R36 (2010).
- Jing, Q., McLellan, A., Greally, J. M. & Suzuki, M. Automated computational analysis of genome-wide DNA methylation profiling data from HELP-tagging assays. *Methods Mol. Biol.* 815, 79–87 (2012).
- Hoffman, M. M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat. Methods 9, 473–476 (2012).
- Kohonen, T. & Somervuo, P. How to make large self-organizing maps for nonvectorial data. Neural Netw. 15, 945–952 (2002).
- Zhang, Y. et al. Chromatin connectivity maps reveal dynamic promoterenhancer long-range associations. Nature 504, 306–310 (2013).
- Geeleher, P. et al. Gene-set analysis is severely biased when applied to genomewide methylation data. Bioinformatics 29, 1851–1857 (2013).
- Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11, R14 (2010).
- Hoyo, C. et al. Association of cord blood methylation fractions at imprinted insulin-like growth factor 2 (IGF2), plasma IGF2, and birth weight. Cancer Causes Control 23, 635–645 (2012).
- Perkins, E. et al. Insulin-like growth factor 2/H19 methylation at birth and risk of overweight and obesity in children. J. Pediatr. 161, 31–39 (2012).
- 36. Godfrey, K. M. et al. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes* 60, 1528–1534 (2011).
- Pettitt, D. & Jovanovic, L. Birth weight as a predictor of type 2 diabetes mellitus: the U-shaped curve. Curr Diab Rep. 1, 78–81 (2001).
- Dashti, M., Peppelenbosch, M. P. & Rezaee, F. Hedgehog signalling as an antagonist of ageing and its associated diseases. *Bioessays* 34, 849–856 (2012).
- Blank, U., Karlsson, G. & Karlsson, S. Signaling pathways governing stem-cell fate. Blood 111, 492–503 (2008).

- Mackie, A. R. et al. Sonic hedgehog-modified human CD34+ cells preserve cardiac function after acute myocardial infarction. Circ. Res. 111, 312–321 (2012).
- Eriksson, J. G., Kajantie, E., Lampl, M., Osmond, C. & Barker, D. J. Markers of biological fitness as predictors of all-cause mortality. *Ann. Med.* 45, 156–161 (2013)
- Ravelli, A. C., van Der Meulen, J. H., Osmond, C., Barker, D. J. & Bleker, O. P. Obesity at the age of 50 y in men and women exposed to famine prenatally. *Am. J. Clin. Nutr.* 70, 811–816 (1999).
- 43. Michels, K. B., Harris, H. R. & Barault, L. Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements. *PLoS One* **6**, e25254 (2011).
- Zhang, F. F. et al. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics 6, 623–629 (2011).
- Perng, W. et al. A prospective study of LINE-1DNA methylation and development of adiposity in school-age children. PLoS One 8, e62587 (2013).
- Murphy, S. K. et al. Gender-specific methylation differences in relation to prenatal exposure to cigarette smoke. Gene 494, 36–43 (2012).
- 47. Pilsner, J. R. et al. Influence of prenatal arsenic exposure and newborn sex on global methylation of cord blood DNA. PLoS One 7, e37147 (2012).
- 48. Khulan, B. et al. Periconceptional maternal micronutrient supplementation is associated with widespread gender related changes in the epigenome: a study of a unique resource in the Gambia. Hum. Mol. Genet. 21, 2086–2101 (2012).
- Avogaro, A., Fadini, G. P., Gallo, A., Pagnin, E. & de Kreutzenberg, S. Endothelial dysfunction in type 2 diabetes mellitus. *Nutr. Metab. Cardiovascular Dis.* 16 (Suppl 1), S39–S45 (2006).
- Asahara, T. et al. Isolation of putative progenitor endothelial cells for angiogenesis. Science 275, 964–967 (1997).
- Zentilin, L. et al. Bone marrow mononuclear cells are recruited to the sites of VEGF-induced neovascularization but are not incorporated into the newly formed vessels. Blood 107, 3546–3554 (2006).
- 52. Ziegelhoeffer, T. et al. Bone marrow-derived cells do not incorporate into the adult growing vasculature. Circ. Res. 94, 230–238 (2004).
- Ianus, A., Holz, G. G., Theise, N. D. & Hussain, M. A. *In vivo* derivation of glucose-competent pancreatic endocrine cells from bone marrow without evidence of cell fusion. *J. Clin. Invest.* 111, 843–850 (2003).
- Lagasse, E. et al. Purified hematopoietic stem cells can differentiate into hepatocytes in vivo. Nat. Med. 6, 1229–1234 (2000).
- Fadini, G. P. et al. Diabetes impairs stem cell and proangiogenic cell mobilization in humans. Diabetes Care 36, 943–949 (2013).
- Loomans, C. J. et al. Endothelial progenitor cell dysfunction: a novel concept in the pathogenesis of vascular complications of type 1 diabetes. *Diabetes* 53, 195–199 (2004).
- Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A. & Weissman, I. L. The aging of hematopoietic stem cells. *Nat. Med.* 2, 1011–1016 (1996).
- Liang, Y., Van Zant, G. & Szilvassy, S. J. Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. *Blood* 106, 1479–1487 (2005).
- Pearce, D. J., Anjos-Afonso, F., Ridler, C. M., Eddaoudi, A. & Bonnet, D. Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging. Stem Cells 25, 828–835 (2007).
- Beerman, I. et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proc. Natl Acad. Sci. USA 107, 5465–5470 (2010).
- Harrison, D. E. Long-term erythropoietic repopulating ability of old, young, and fetal stem cells. J. Exp. Med. 157, 1496–1504 (1983).
- Pranke, P. et al. Hematologic and immunophenotypic characterization of human umbilical cord blood. Acta. Haematol. 105, 71–76 (2001).
- Suzuki, M. & Greally, J. M. DNA methylation profiling using HpaII tiny fragment enrichment by ligation-mediated PCR (HELP). *Methods* 52, 218–222 (2010).
- Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics 10, 232 (2009).
- Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005).

Acknowledgements

Support for this project was provided by the Roadmap Epigenomics Program, R01 HD063791 (Einstein/Greally). Support was also provided by Einstein's Medical Student Training Program (to N.A.W. and J.N.T., NIH NIGMS T32 GM007288), the Reproductive Scientist Development Program (to H.J.H., NICHD K12HD00849-26) and Einstein's Center for Epigenomics, including the Epigenomics Shared Facility and Computational Epigenomics Group.

Author contributions

F.D., N.A.W., J.N.T. and H.J.H.: performed the experiments and analysed data. Y.M.Z.: performed the experiments. N.A.W. and J.M.G.: guided and performed analytical approaches. F.D.: wrote the manuscript. F.H.E. and J.M.G.: designed the study, analysed data and wrote the manuscript.

Additional information

Accession codes: HELP-Tagging data have been deposited in the Gene Expression Omnibus (GEO) database under the accession code GSE53177.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

 $\label{lem:compression} \textbf{Reprints and permission} \ information \ is available \ online \ at \ http://npg.nature.com/reprintsandpermissions/$

How to cite this article: Delahaye, F. *et al.* Sexual dimorphism in epigenomic responses of stem cells to extreme fetal growth. *Nat. Commun.* 5:5187 doi: 10.1038/ncomms6187 (2014)