Exposure to aflatoxin B_1 *in utero* is associated with DNA methylation in white blood cells of infants in The Gambia

Hector Hernandez-Vargas[1], Jovita Castelino[2], Matt J Silver[3], Paula Dominguez-Salas[3], Marie-Pierre Cros[1], Geoffroy Durand[4], Florence Le Calvez-Kelm[4], Andrew M Prentice[3], Christopher P Wild[5], Sophie E Moore[3,6], Branwen J Hennig[3], Zdenko Herceg[1], Yun Yun Gong[7*] and Michael N Routledge[2*]

[1]Epigenetics Group, International Agency for Research on Cancer (IARC), Lyon, France

[2]LICAMM, School of Medicine, University of Leeds, Leeds, UK

[3]MRC International Nutrition Group at LSHTM, UK & MRC Keneba, MRC Unit, The Gambia

[4] Genetic Cancer Susceptibility Group, IARC, Lyon, France

[5] Director, IARC, Lyon , France

[6]MRC Human Nutrition Research, Cambridge

[7]UK Institute of Global Food Security, Queen's University Belfast, Belfast, UK

Corresponding author:

Dr Michael N Routledge, LICAMM, LIGHT Building, University of Leeds, Leeds, LS2 9JT, UK.

Tel: +44(0)1133437763

E-mail: medmnr@leeds.ac.uk

*These authors contributed equally

Summary

Background: Exposure to environmental toxins during embryonic development may lead to epigenetic changes that influence disease risk in later life. Aflatoxin is a contaminant of staple foods in sub-Saharan Africa, is a known human liver carcinogen and has been associated with stunting in infants.

Methods: We have measured aflatoxin exposure in 115 pregnant women in The Gambia and examined the DNA methylation status of white blood cells from their infants at two to eight months old (mean 3·6±0·9). Aflatoxin exposure in women was assessed using an ELISA method to measure aflatoxin albumin adducts in plasma taken at 1-16 weeks of pregnancy. Genome-wide DNA methylation of infant white blood cells was measured using the Illumina Infinium HumanMethylation450 beadchip.

Results: AF-alb levels ranged from 3·9 to 458·4 pg/mg albumin. We found that aflatoxin exposure in the mothers was significantly associated to DNA methylation in their infants for 71 CpG sites (FDR<0.05), with an average effect size of 1.7% change in methylation. Aflatoxin-associated differential methylation was observed in growth factor genes such as *FGF12* and *IGF1*, and immune related genes such as *CCL28*, *TLR2*, and *TGFBI*. Moreover, one aflatoxin-associated methylation region (corresponding to the miR-4520b locus) was identified.

Conclusions: This study shows that maternal exposure to aflatoxin during the early stages of pregnancy is associated with differential DNA methylation patterns of infants, including in genes related to growth and immune function. This reinforces the need for interventions to reduce aflatoxin exposure, especially during critical periods of foetal and infant development.

Keywords: Aflatoxin, DNA methylation, in utero exposure

Introduction

It has been proposed that changes to the epigenome during foetal development can contribute to disease susceptibility in adulthood.[1,2] Critical developmental periods exist during which the foetus is sensitive to the environment and adapts accordingly to prepare for survival following birth.[3] Additionally, the earliest point of embryogenesis is a time of marked epigenetic change wherein genome-wide demethylation of the oocyte and sperm genomes occurs, followed by *de novo* genome-wide and tissue-specific methylation.[4] During this period, environmental exposures and stresses can influence the developing epigenome, causing life-long phenotypic alterations and potentially resulting in increased susceptibility to adult disease. Environmental exposure-specific patterns of DNA methylation could thus act as biomarkers with potential predictive and prognostic value.

Aflatoxins are secondary metabolites produced by certain strains of *Aspergillus* fungi, and have been classified as a Group I carcinogen by the International Agency for Research on Cancer (IARC).[5] Exposure to aflatoxins, of which aflatoxin B_1 (AFB_1) is the most common and most toxic, primarily occurs through the consumption of contaminated maize and groundnuts, which form the basis of staple diets in many low- and middle income countries (LMIC). Exposure to AFB_1 can occur *in utero* given that the toxin can cross the placental barrier.[6] Altered methylation at several genes has been observed in the context of AFB_1 exposure and liver cancer. Hypermethylation of the *GSTP1* gene promoter has been associated with levels of AFB_1 DNA adducts in hepatocellular carcinoma (HCC) tumour tissue in Taiwanese patients[7] and *RASSF1A* hypermethylation has been associated with AFB_1 DNA adducts in HCC in SE China.[8] Decreased *LINE1* and *SAT2* DNA methylation has been associated with aflatoxin exposure of both HCC patients and healthy adults in Taiwan.[9]

In this report, we studied the consequences of early life exposure to aflatoxin at the DNA methylation level using a hypothesis-free genome-wide approach. The methylome-wide

analysis of infant's DNA from a Gambian cohort reveals non-random differences in methylation related to early-life aflatoxin exposure.

Methods

Sample selection and preparation

Ethical approval was obtained from the joint Gambia Government/MRC Unit, The Gambia Ethics Committee. Pregnant women and later their infants were recruited in the West Kiang region of The Gambia; details of this cohort have been described.[10–12] Pregnant women aged 18–45y provided a blood sample at 1–16 weeks of pregnancy for biochemical analyses. Date of conception was estimated by an obstetric ultrasound examination at the point women booked for antenatal care. A total of 115 infants between 2–8 months of age (mean 3·6 ± SD 0·9) provided a blood sample for DNA extraction. Only four of the children were over 6 months old when blood was taken.

Assessment of aflatoxin exposure

Plasma derived from blood obtained from these women were analysed for aflatoxin-albumin adduct (AF-alb) levels by a competitive ELISA as described previously.[10,13] Samples were analysed in triplicate, repeated on at least two separate days. Intra-assay coefficient of variation (CV) was < 15 % and inter-assay CV was ≤ 25 %. Three positive quality controls of different known AF-alb levels and one negative control were analysed with each batch of samples. AF-alb levels are presented with reference to the amount of albumin in the sample (i.e. pg AF-alb/mg albumin). The limit of detection for AF-alb was 0·6 pg/mg albumin.

An aliquot (20 µl) from each plasma sample was taken to measure albumin levels using a commercial kit (Bio-quant, San Diego, CA, USA) based on the bromocresol green (BCG) method.

Bisulfite conversion and DNA methylation assessment

Blood (3 ml) collected from the infants at 2-8 months of age was used for DNA extraction from white blood cells (WBC) as previously described.[11] The extracted DNA (500 ng) was bisulfite modified using the EZ DNA Methylation kit (Zymo Research, D5001) following the manufacturer's instructions for Illumina Infinium HumanMethylation450 (HM450) beadchip assay. Modified DNA was stored at -20˚C until needed. To quantify the percentage of methylated cytosine in individual CpG sites, we performed bisulfite pyrosequencing, as previously described.[14]

Genome-wide methylation profiles were obtained using the HM450 Infinium methylation bead arrays (Illumina, San Diego, USA). Briefly the HM450 beadchip interrogates more than 480,000 methylation sites.[15] The analysis on the bead array was conducted following the recommended protocols for amplification, labelling, hybridization and scanning.

Bioinformatic analyses

Data pre-processing and analysis was performed using R/Bioconductor packages. Data quality was assessed using boxplots for the distribution of methylated and unmethylated signals, and multidimensional scaling plots and unsupervised clustering were used to check for sample outliers. Quantile normalized data were used to infer blood cell subtypes based on Houseman's regression calibration, as previously described.[16,17] Cross-reactive probes, probes mapping to sex chromosomes, and probes overlapping with a known single nucleotide polymorphism (SNPs) with an allele frequency of at least 5% in the overall population (all ethnic groups) were also removed, as previously described.[18] After background correction and colour-bias adjustment, type I and type II probe distributions were aligned using the intra-array beta-mixture quantile normalization from the watermelon package.[19,20] The Beta-value is the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities). Although Beta-values are widely used for

data interpretation, its logarithmic transformation (M-value) has been shown to perform better in some downstream analyses [21]. The M-value was calculated as the log2 ratio of the intensities of methylated probe versus unmethylated probe. After batch correction (sva package),[22] M-values were interrogated for association with aflatoxin exposure (AFB-associated loci), by modelling the study variables and covariates (i.e. aflatoxin exposure, infant sex, and season of conception) together with latent surrogate variables in a linear regression using the limma package.[23] In the initial analysis, aflatoxin exposure was studied as a categorical variable by dividing the samples into "Low" and "High" exposure groups, based on the median of 33·2 pg/mg as cut-off. For all further analyses, aflatoxin exposure data was used as a continuous variable. AFB-associated methylation sites (AfMSs) were selected based on a threshold for the adjusted P value (False Discovery Rate or FDR) of 0·05. For pathway/ontology analyses we performed a Bonferroni correction of the raw P values to adjust for the number of probes in the corresponding gene. Then, for each gene we selected the probe with the minimum Bonferroni-corrected P, and P values were further adjusted for the number of gene symbols on the array. Those genes with an FDR-adjusted $P < 0.05$ were taken for further pathway analyses, using the Enrichr gene list enrichment web tool.[24] The bump hunting method (minfi package) was used to define AFB-associated regions using the recommended proximity-based criteria.[25] An Aflatoxin-associated methylation region (AfMR) was defined by the presence of at least 2 differentially methylated CpG sites with a maximum gap of 500 bp.The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE59592.

Results

Differential methylation associated with aflatoxin exposure in Gambian infants

A description of the infant samples and covariates is presented in Table 1, together with aflatoxin-albumin (AF-alb) adduct levels in maternal blood obtained during the first trimester of pregnancy (Figure 1a). DNA methylation data was available for 115 infants across 451041 sites. In the initial analysis characterizing the effect of early life Low vs High aflatoxin exposure on the DNA methylome, no significant differential methylation was observed at FDR < 0·05. A selection of top CpG sites (lowest p values) with at least 3% difference in methylation between Low and High groups was correctly validated by pyrosequencing (Figure S1).

In a second analysis, the association of both single-locus and regional DNA methylation with AF-alb modelled as a continuous variable was assessed, considering the continuous distribution of the AF-alb data (Figure 1a). Also in this analysis we tested the association between AF-alb adduct levels (n=115) and the methylation of 451041 cytosines. A set of 71 CpG sites were significantly correlated with AF-alb (FDR < 0·05, R-squared = 0.88 and adjusted R-squared = 0.859 for the overall analysis), and were defined as aflatoxin-associated methylation sites (AfMSs) (Table 2). The average absolute difference across the 71 CpGs was 0.017 (1.7%), and a quantile-quantile plot suggests no systematic inflation of P values (Figure 1b). AfMSs were not enriched in specific chromosomal locations (Figure 1c) and their distribution in relation to genes and CpG islands (CGIs) followed the proportions of the total probe content of the methylation array, mapping to gene bodies and CGIs, respectively (Figure 1d). AfMSs correlated either positively or negatively with aflatoxin exposure (Figure 1e and 1f). Results were similar when including additional covariates in the regression model (i.e. maternal age, maternal body mass index (BMI); data not shown). Similarly, analyses including age of sample collection as a covariate in the regression model yielded similar results (86 AfMSs, 62 of them in common with the first analysis). To rule out the possibility that results could be driven by few outliers with extreme DNA methylation values, we

performed several analyses after removing one or more samples. Removing up to 4 samples with extreme values for the top AfMSs reduced the statistical significance of the analyses, possibly due to the reduced sample size. However, the top significant loci remained essentially unchanged in all cases. Finally, region-level analyses showed only one aflatoxin-associated methylation region (AfMR) (adjusted p value < 0·05) containing 5 informative CpG sites at the microRNA *hsa-miR-4520B* locus.

Importantly, aflatoxin-associated loci included growth factors genes such as *FGF12* and *IGF1*, and immune related genes such as *CCL28, TLR2*, and *TGFBI*. In addition, AfMSs included three sites within the TRNA-YW (Wybutosine) Synthesizing Homolog gene *TYW3* locus. However, performing pathway/ontology analysis in HM450 data may result in spurious associations due to the unbalanced representation of probes for different genes within the array.[26] To overcome this issue, we adjusted for the number of probes per gene symbol and selected only those genes with at least one significant CpG site below the FDR-ajusted P value threshold of 0.05 (see Methods). Using this strategy, we obtained 53 unique gene symbols for pathway analysis (out of 71 AfMSs on the original analysis). Significant pathways (FDR < 0.05 and containing at least 3 genes from the list) included microRNA 186 (miR-186) using TargetScan (targeting 5 genes from the list: *RFWD2, PAN3, CLDN1, RNF11* and *CXXC5*), the Transcription Factor Protein-Protein Interactions Vitamin D Receptor (*ZBTB16, CXXC5,* and *PRKCSH*), the GEO Kinase perturbations term Anaplastic Lymphoma Receptor tyrosine kinase (ALK) (25 targets found), the GO positive regulation of Ras GTPase activity (GO:0032320) (*SERGEF, DOCK2, TBC1D28*), and the Human Phenotype Ontology term Hyperbilirubinemia (*HK1, PRKCSH, SPTB*).

Leukocyte-adjusted methylome analyses

DNA methylation is frequently tissue- and cell-type specific, and analyses may therefore be confounded in samples containing multiple cell types. DNA methylation at loci known to be differentially methylated across cell types have been used to quantify multiple cell types in complex mixtures.[27] Using this recently available algorithm we found no differences in the inferred proportions of six cell subtypes (B cells, NK cells, monocytes, CD8 T cells, CD4 T cells, and granulocytes), when comparing the two categories of aflatoxin exposure "Low" and "High" (Figure 2a). However, correlation analyses showed that increasing maternal exposure to aflatoxin significantly correlated with a reduced proportion of CD8 T cells (P= 0·0419, Spearman= -0·19) and an increased percentage of granulocytes (P= 0·0423, Spearman= 0·189) (Figure 2b). Therefore, to rule out a confounding effect of cell subtypes on the aflatoxin-methylation analysis, the inferred proportion of the six cell subtypes was included in a new regression model to test for AfMSs. Ninety-one AfMSs were obtained with an FDR<0·05, out of which 67 were present in the 71 CpG list when cell subtypes were not included in the model.

The inferred differences in blood cell subpopulations may be the result of biological effects of aflatoxin exposure. However, differential methylation associated with aflatoxin exposure does not seem to be highly influenced by the proportion of cell subpopulations.

Discussion

DNA methylation is one type of epigenetic marker that may be modulated by interaction with environmental factors. Here we have shown for the first time that dietary exposure of pregnant women to aflatoxin is associated with genome-wide DNA methylation in the WBC of their infants.

Hypermethylation of specific genes has been observed in a number of human tumours, and is a mechanism by which genes such as tumour suppressor genes may be inactivated during carcinogenesis.[28] There is increasing evidence that exposure to environmental toxins results in altered DNA methylation in not only tumours but in a range of normal tissues as well.[29] Where this occurs following exposure *in utero*, it is possible that changes may be linked to subsequent adverse health outcomes,[2] although to date human studies have been few.

As well as being a known liver carcinogen, aflatoxin has been associated with child growth impairment,[30,31] including from exposure that occurred *in utero*.[6] The mechanism by which aflatoxin exerts its effects on child growth is not clear, but epigenetic changes occurring during embryonic development could play a role. The women and children who took part in this study live in a region of Gambia where, as in many other regions in Africa, aflatoxin exposure is prevalent due to the fungal contamination of staple crops such as groundnuts.[10] As a result, children are exposed to aflatoxin throughout their childhood and later life, and it is likely that exposure *in utero* contributes to health impacts in childhood and later. In a population such as this where groundnut consumption is a staple part of the diet, individual exposure to aflatoxin depends largely on the levels of contamination of the groundnuts rather than variation in groundnut consumption (with aflatoxin being heterogeneously distributed in groundnuts).

Aflatoxin exposure in the pregnant women was assessed using the well-validated AF-alb biomarker in blood, which provides a reliable method for measuring differences in aflatoxin exposure.[32] As albumin has a serum half life of about 20 days, measuring AF-alb integrates exposure that has occurred over a period of time prior to the sampling, which reduces any error associated with assessing exposure during early pregnancy from single sampling.

Methylation at 71 CpG sites was significantly correlated with aflatoxin exposure. Fifty-two of the 71 sites are located in annotated genes, including a number that are involved in the immune response or the inflammatory response (eg *TLR2* and *CCL28*). This is of interest because compromised defences against infection have been hypothesised as one explanation for how aflatoxin exposure leads to growth inhibition.[33] The aflatoxin-related differences in absolute DNA methylation were typically small. For the 71 AfMPs we compared the mean of DNA methylation between those subjects not exposed (<10 pg/mg AF-albumin) and those highly exposed (>100 pg/mg AF-albumin). The average absolute difference across the 71 CpGs was 0.017 (1.7%). The biological relevance of this small change in methylation is not known. However, this is similar to what has been reported in other population-based studies using healthy subjects. For example, it was recently shown that DNA methylation of 353 CpG sites is able to predict chronological age with remarkable accuracy, even though the absolute difference across all CpG sites was only 0·032.[34] Although there are a number of reports on small significant differences in methylation, especially on population-based studies, the biological relevance of these small differences is uncertain. Until replication and further validation can be done, we only rely on biological plausibility to evaluate possible functional relevance.

Previous studies have shown that environmental exposures including dietary folate, smoking and constituents of air pollution are associated with altered DNA methylation profiles in WBC.[35–37] In the population studied here, season of conception has previously been shown to influence methylation of metastable epialleles in WBC of children,[11] but season of conception was not a confounder for the aflatoxin-associated levels of genome-wide DNA methylation observed here. The influence of environmental exposures on DNA methylation during pregnancy has been explored in a number of recent methylome-wide studies. Cadmium exposure during pregnancy in a cohort of women from a polluted region of

Bangladesh was associated with DNA methylation differences in cord blood, with sex-specific levels of DNA methylation being observed.[38] Cadmium related DNA methylation was also associated with lower birth weight. Koestler et al[39] found an association between maternal arsenic exposure during pregnancy and differences in DNA methylation measured in cord blood of infants from New Hampshire, USA. Cigarette smoking during pregnancy has also been demonstrated to alter DNA methylation in specific loci.[37] Interestingly, changes in DNA methylation in cord blood that were associated with pre-pregnancy BMI of the mothers have been found to persist in DNA of children at age 3 years,[41] so such changes can be long lasting, with the potential for long-term effects. Most recently, exposure to arsenic during early pregnancy has been found to be associated with decreased methylation in cord blood, with a sex-specific pattern being observed in that study as decreased methylation was more pronounced in boys.[42] Whilst exposure to aflatoxin has been shown to be associated with *LINE1* and *SAT2* methylation in adults,[9] our methylome-wide study has shown that exposure to aflatoxin at a critical period during early development modulates DNA methylation in a set of protein coding genes and is the first time that exposure to aflatoxin has been associated with DNA methylation differences in children. Although we cannot rule out that blood DNA methylation was influenced by additional exposures during post-natal life, most samples in our study were collected before the children were 6 months old, and before this age aflatoxin exposure is known to be very low.[31] Other limitations of our study include the possibility of false positive results, the cellular complexity inherent to blood samples, and the small magnitude of the differences in DNA methylation. Therefore, replication in a larger cohort will be necessary to further validate an association between DNA methylation and aflatoxin exposure.

In summary, aflatoxin exposure during pregnancy associates to differential methylation in infant's DNA at specific loci. Our findings that exposure to aflatoxin *in utero* is associated

with DNA methylation patterns across a number of genes at age 2–8 months may be relevant to the mechanism of aflatoxin-related child stunting, or liver cancer in later life. These biological effects suggest potential avenues for research into the mechanism by which aflatoxin influences child growth and other health outcomes.

Funding

Sample collection in The Gambia was supported by core funding MC-A760-5QX00 to the International Nutrition Group by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and by a Wellcome Trust grant (WT086369MA). YYG, MNR and JMC were supported by a grant from the International Agency for Research on Cancer. ZH and the Epigenetics Group at IARC is supported by a Grand Challenges Exploration Grant from the Bill and Melinda Gates Foundation (Grant Number: OPP1061062). YYG and CPW were also supported by the National Institute of Environmental Health Sciences (grant ES06052).

Acknowledgements

This study depended on the participation of the women and children of West Kiang, to whom we are grateful. We also acknowledge the assistance of field workers, nurses and technicians at MRC Keneba.

Conflict of interest: The authors declare they have no actual or potential competing financial interests.

Tables

Table 1: Characteristics of study participants

	High aflatoxin exposure group		Low aflatoxin exposure group		Total	
	N.	**Mean ± SD (min-max)**	**N.**	**Mean ± SD (min-max)**	**N.**	**Mean ± SD (min-max)**
Male % (total number of infants)		47.2% (55)		56·7% (60)		52.1% (115)
Age (months)	51	3.7 ± 1.0 (2·1–7·3)	56	3·5 ± 0.7 (2.6–7·9)	107[a]	3·6 ± 0·9 (2·1–7.9)
Maternal BMI (kg/m^2)	55	21·0 ± 3·1 (15·3–33·9)	60	20·6 ± 3·00 (15·1–30·1)	115	20·8 ±3·0 (15·1–33·9)
Maternal age (years)	55	28.9 ± 6·46 (17·9–43·1)	60	29.21 ± 6·9 (17·5–45·5)	115	29·1 ± 6·7 (17·5–45·5)
AF-alb (pg/mg albumin) Dry season (November to May)[b]	30	75·83 (58·75, 97·89)	27	13·85 (11·29, 16·99)**	57	34·36 (26·05, 45·32)
AF-alb (pg/mg albumin) Rainy season (June to October)[a]	25	78·96 (60·85, 102·46)	33	18·20 (15·44, 21·46)**	58	35·23 (27·76, 44·70)

[a] child age was missing from eight datasets

[b] AF-alb adduct levels presented as geometric means with 95% confidence intervals (95% CI) in parentheses.

** $P < 0·001$ for means between adjacent groups

Table 2: Aflatoxin-associated methylation sites (AfMSs) in all infants.

TargetID	nearest Gene	logFC	P value	FDR	distance to gene (bp)	Description	No. probes in HM450
cg16035199	ODF3L2	0.008	7E-09	8E-04	0	outer dense fiber of sperm tails 3-like 2	22
cg12251659	KDM2B	0.0131	2E-09	8E-04	0	lysine (K)-specific demethylase 2B	82
cg12940473	TLR2	-0.007	2E-08	0.001	0	toll-like receptor 2	12
cg24402300	EPS8L1	-0.01	1E-08	8E-04	0	EPS8-like 1	22
cg11482794	TGFBI	-0.008	1E-08	8E-04	0	transforming growth factor, beta-induced, 68kDa	22
cg03521258	LSR	-0.009	2E-08	0.001	611	lipolysis stimulated lipoprotein receptor	14
cg16209795	FGF12	-0.008	7E-09	8E-04	0	fibroblast growth factor 12	52
cg09524880	PRKCSH	-0.003	6E-08	0.002	0	protein kinase C substrate 80K-H	6
cg26855724	CRYZ	0.0036	4E-07	0.008	0	crystallin, zeta (quinone reductase)	1
cg24881229	TAF1D	-0.006	5E-07	0.008	0	TATA box binding protein (TBP)-associated factor, RNA polymerase I, D, 41kDa	1
cg19499998	SLC22A2	-0.003	2E-08	0.001	249	solute carrier family 22 (organic cation transporter), member 2	21
cg00121533	TYW3	0.0039	3E-07	0.007	0	tRNA-yW synthesizing protein 3 homolog (S. cerevisiae)	2
cg16779839	BAHCC1	0.0125	6E-09	8E-04	0	BAH domain and coiled-coil containing 1	112
cg02115904	CCDC90B	-0.007	5E-08	0.002	0	coiled-coil domain containing 90B	15
cg10832470	CBY3	-0.006	8E-08	0.003	267	chibby homolog 3 (Drosophila)	9
cg21963925	CACNA1H	0.0064	1E-08	8E-04	0	calcium channel, voltage-dependent, T type, alpha 1H subunit	162
cg14154441	PAN3	0.0084	3E-07	0.007	0	PAN3 poly(A) specific ribonuclease subunit	9
cg09563228	GAL	0.0062	2E-07	0.005	9846	galanin/GMAP prepropeptide	14
cg23581186	KRTAP21-2	-0.004	4E-06	0.034	0	keratin associated protein 21-2	1
cg21240861	DNAJB6	-0.006	5E-08	0.002	56	DnaJ (Hsp40) homolog, subfamily B, member 6	89
cg21535942	TYW3	0.0022	3E-06	0.026	0	tRNA-yW synthesizing protein 3 homolog (S. cerevisiae)	2
cg10091686	LOC729732	0.0023	6E-06	0.045	0	unknown	unknown
cg21919136	CLDN1	0.0023	6E-07	0.009	7	claudin 1	13
cg00063979	CXXC5	-0.006	2E-07	0.004	12605	CXXC finger protein 5	50
cg21796825	TMEM5	-0.002	5E-07	0.008	0	transmembrane protein 5	15
cg04767522	KIF13A	0.0094	2E-07	0.005	0	kinesin family member 13A	37
cg26064460	TCEA3	0.0099	5E-07	0.008	3625	transcription elongation factor A (SII), 3	19
cg25741192	RFWD2	0.0042	6E-07	0.009	550	ring finger and WD repeat domain 2, E3 ubiquitin protein ligase	16
cg26889819	PLEKHG4B	-0.005	2E-07	0.005	0	pleckstrin homology domain containing, family G (with RhoGef domain) member 4B	47
cg20540566	TTPA	-0.006	8E-07	0.01	0	tocopherol (alpha) transfer protein	14
cg02570501	ZNF107	0.0091	8E-07	0.01	952	zinc finger protein 107	14
cg05419696	CYB5D1	-0.005	6E-06	0.043	0	cytochrome b5 domain containing 1	2
cg17325789	EBF3	0.0037	8E-08	0.003	1040	early B-cell factor 3	162

cg10752421	SLC7A1	0.0047	4E-07	0.008	0	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	36
cg05379509	SPTB	-0.004	5E-07	0.008	0	spectrin, beta, erythrocytic	32
cg11515089	TBC1D28	0.0087	3E-06	0.03	5792	TBC1 domain family, member 28	5
cg07425780	GPRC5B	0.0037	8E-07	0.01	0	G protein-coupled receptor, class C, group 5, member B	21
cg01733928	KCNK10	-0.013	4E-07	0.008	21052	potassium channel, two pore domain subfamily K, member 10	42
cg15733917	LIMS1	-0.013	3E-06	0.03	115	LIM and senescent cell antigen-like domains 1	7
cg21536096	HAND2	0.0039	3E-06	0.029	29665	heart and neural crest derivatives expressed 2	10
cg14294658	SERGEF	-0.002	1E-06	0.012	0	secretion regulating guanine nucleotide exchange factor	38
cg26321643	HES4	-0.003	3E-06	0.026	1361	hes family bHLH transcription factor 4	16
cg06600936	ALPK3	-0.003	3E-06	0.026	0	alpha-kinase 3	17
cg26680885	RNASET2	-0.004	4E-06	0.031	0	ribonuclease T2	13
cg17079961	PRKAB2	-0.002	3E-06	0.03	239	protein kinase, AMP-activated, beta 2 non-catalytic subunit	15
cg05151395	SLC25A35	0.0014	7E-06	0.045	0	solute carrier family 25, member 35	8
cg03784882	AMZ1	-0.003	1E-06	0.015	3174	archaelysin family metallopeptidase 1	41
cg03474133	CCNI	-0.003	4E-06	0.035	386	cyclin I	13
cg15813673	ZBTB16	0.0059	1E-06	0.012	0	zinc finger and BTB domain containing 16	54
cg04851471	DOCK2	-0.005	2E-06	0.022	0	dedicator of cytokinesis 2	32
cg16527041	SPATA5	0.0057	4E-06	0.035	0	spermatogenesis associated 5	15
cg00101118	RNF11	-0.004	5E-06	0.04	203	ring finger protein 11	14
cg18199554	CCL28	0.0046	5E-06	0.038	0	chemokine (C-C motif) ligand 28	16
cg01476003	ZNF23	0.0021	6E-06	0.043	21015	zinc finger protein 23	14
cg01511465	MC1R	0.0054	7E-06	0.045	0	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)	14
cg05239504	HERPUD2	0.0053	5E-06	0.036	20898	HERPUD family member 2	23
cg16001913	HK1	0.0029	2E-06	0.025	110	hexokinase 1	47
cg23666856	TSSC1	-0.006	1E-06	0.012	0	tumor suppressing subtransferable candidate 1	118
cg15241635	AGAP1	0.0029	6E-07	0.009	0	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1	297
cg02893344	GSE1	0.0032	2E-06	0.018	127758	Gse1 coiled-coil protein	121
cg18395623	SCRN1	0.0027	7E-06	0.045	0	secernin 1	28
cg04351541	KIF1A	-0.002	4E-06	0.033	0	kinesin family member 1A	52
cg05487269	FLYWCH1	-0.002	7E-06	0.045	0	FLYWCH-type zinc finger 1	34
cg08005809	SCRIB	0.0048	6E-06	0.043	0	scribbled planar cell polarity protein	46
cg16419361	AGAP3	0.0023	7E-06	0.047	0	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3	54
cg01248385	MCF2L	0.0057	2E-06	0.018	0	MCF.2 cell line derived transforming sequence-like	285
cg02613818	IGF1R	-0.008	7E-06	0.045	0	insulin-like growth factor 1 receptor	132
cg19547192	PTPRN2	-0.006	8E-07	0.01	0	protein tyrosine phosphatase, receptor type, N polypeptide 2	1210
cg09848638	PRKAR1B	-0.003	5E-06	0.04	0	protein kinase, cAMP-dependent, regulatory, type I, beta	205
cg05361818	TNXB	-0.004	6E-06	0.042	0	tenascin XB	509
cg15689733	PTPRN2	0.002	3E-06	0.026	0	protein tyrosine phosphatase, receptor type, N polypeptide 2	1210

logFC: log2 fold change; FDR: false discovery rate; analysis adjusted for infant sex and season of conception; n=115.

References

1. Waterland RA, Michels KB. Epigenetic epidemiology of the developmental origins hypothesis. *Ann Rev Nutr* 2007; **27**: 363-88.
2. Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat Rev Endocrinol* 2009; **5**, 401-8.
3. Barker D. The developmental origins of adult disease. *J Am Coll Nutr* 2004; **23**: 588S-95S.Bird A. DNA methylation patterns and epigenetic memory. *Gene Dev* 2002; **16**: 6-21.
4. IARC. Some traditional herbal medicines, some mycotoxins, naphthalene and styrene. *IARC monographs on the evaluation of carcinogenic risks to humans / World Health Organization, International Agency for Research on Cancer* 2002;**82**:1-556.
5. Turner PC, Collinson AC, Cheung YC et al. Aflatoxin exposure *in utero* causes growth faltering in Gambian infants. *Int J Epidemiol* 2007;**36**:1119-25.
6. Zhang YJ, Chen Y, Ahsan H, et al. Silencing of glutathione S-transferase P1 by promoter hypermethylation and its relationship to environmental chemical carcinogens in hepatocellular carcinoma. *Cancer Lett* 2005; **221**:135-43.
7. Feng Y, Xue WJ, Li P et al. (2012) RASSF1A hypermethylation is associated with aflatoxin B1 and polycyclic aromatic hydrocarbon exposure in hepatocellular carcinoma. *Hepato-Gastroenterol* 2012;**59**:1883-88.
8. Wu HC, Wang Q, Yang HI, Tsai WY, Chen CJ and Santella RM. Global DNA methylation in a population with aflatoxin B1 exposure. *Epigenetics* 2013;**8**:962-69.
9. Castelino JM, Dominuez-Salas P, Routledge MN et al. Seasonal and gestation stage associated differences in aflatoxin exposure in pregnant Gambian women. *Trop Med Int Health* 2014;**19:**348-54.
10. Dominguez-Salas P, Moore SE, Baker MS et al. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nature Comms* 2014;**5**:3746.
11. Moore SE, Fulford AJ, Darboe MK et al. 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* 2012;**12**:107.
12. Chapot B, Wild CP. 1991. ELISA for quantification of aflatoxin-albumin adduct and their application to human exposure assessment. In: Warhol M, van Velzen D, Bullock GR (eds). Techniques in Diagnostic Pathology. Academic Press, San Diego CA, 1991:135-55.
13. Hernandez-Vargas H, Lambert MP, Le Calvez-kelm F et al. Hepatocellular carcinoma displays DNA methylation signatures with potential as clinical predictors. *PLoS One* 2010;**5**: e9749.
14. Bibikova M, barnes B, Tsan C et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011;**98**:288-95.
15. Jaffe AE, Irizarry RA Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014;**15**:R31.
16. Houseman EA, Accomando WP, Koestler DC et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;**13**:article no. 86.

17. Chen YA, Lemire M, Choufani S et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium Human Methylation 450 microarray. *Epigenetics* 2013;**8**: 203-9.
18. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 2013;**14**:293.
19. Teschendorff AE, Marabita F, Lechner M et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. *Bioinformatics* 2013;**29**:189-96.
20. Du P, Zhang XA, Huang CC et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2012:article no. 587.
21. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**:882-3.
22. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol* 2004;**3**:article no. 3.
23. Chen EY, Tan CM, Kou Y et al. Enrichr:interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013;**14**:128.
24. Jaffe AE, Murakami P, Lee H et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012;**41**:200-9.
25. Harper KN, Peters BA, Gamble MV. Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. *Cancer Epi Biom Prev* 2013;**22**:1052-60.
26. Bodreddigari S, Jones LK, Egner PA et al. Protection against aflatoxin B_1-induced cytotoxicity by expression of the cloned aflatoxin B_1-aldehyde reductases rat AKR7A1 and human AKR7A3. *Chem Res Toxicol* 2008;**21**: 1143-42.
27. Accomando WP, Wienke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol* 2014;**15**:article no. R50.
28. Herceg Z, Vaissiere T. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics* 2011;**6**:804-19.
29. Cortessis VK, Thomas DC, Levine AJ et al. Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. *Hum Genet* 2012;**131**:1565-89.
30. Gong YY, Cardwell K, Hounsa A et al. Dietary aflatoxin exposure and impaired growth in young children from Benin and Togo: cross sectional study. *Br Med J* 2002;**325:**20-2.
31. Gong YY, Hounsa A, Egal S et al. Postweaning exposure to aflatoxin results in impaired child growth: A longitudinal study in Benin, West Africa. *Environ Health Persp* 2004;**112**:1334-38.
32. Routledge MN, Gong YY. Developing biomarkers of human exposure to mycotoxins. (Ed) S De Saeger, In *Determining mycotoxins and mycotoxigenic fungi in food and feed*, Woodhead Publishing, Cambridge, UK, 2011:225-44.
33. Gong YY, Turner PC, Hall AJ, Wild CP. Aflatoxin exposure and impaired child growth in West Africa: an unexplored international public health burden? In (Eds) JF Leslie, R Badyopadhyay & A Visconti, *Mycotoxins: Detection methods, management, public health and agricultural trade*, CABI, Wallingford, 2008:53-66.
34. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013:**14**;R115.

35. Zhang FF, Santella RM, Wolff M, Kappil MA, Markowitz SB, Morabia A. White blood cell global methylation and *IL-6* promoter methylation in association with diet and lifestyle risk factors in a cancer-free population. *Epigenetics* 2012;**7**: 606-14.
36. Maccani JZJ, Koestler DC, Houseman EA, Marsit CJ, Kelsey KT. Placental DNA methylation alterations associated with maternal tobacco smoking at the *RUNX3* gene are associated with gestational age. *Epigenomics* 2013;**6**:619-30.
37. De Prins S, Koppen G, Jacobs G et al. Influence of ambient air pollution on global DNA methylation in healthy adults: A seasonal follow-up. *Environ Int* 2013;**59**:418-24.
38. Kippler M, Engström K, Mlakar SJ et al. Sex-specific effects of early life cadmium exposure in DNA methylation and implications for birth weight. *Epigenetics* 2013;**8**: 494-503.
39. Koestler DC, Avissar-Whiting M, Houseman EA, Karagas MR, Marsit CJ. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic *in utero*. *Environ Health Persp* 2013;**121**:971-7.
40. Herbstman JB, Wang S, Perera FP et al. Predictors and consequences of global DNA methylation in cord blood and at three years. *PLoS One* 2013;**8**:e72824.
41. Broberg K, Ahmed S, Engström K et al. Arsenic exposure in early pregnancy alters genome-wide DNA methylation in cord blood, particularly in boys. *J Dev Orig Hlth Dis* 2014;**5**:288-98.

Figure Legends

Figure 1. Differential methylation associated with early life exposure to aflatoxin. a. Distribution of aflatoxin exposure during pregnancy in all mothers (left panel) or according to the infant's sex (right panel). **b.** Quantile-quantile plot of the p values after the association between DNA methylation and aflatoxin exposure (as a continuous variable). **c.** Manhattan plot to illustrate the distribution of p values across somatic chromosomes. **d.** Distribution of AfMSs relative to CpG islands (CGI) (i.e. islands, shores, shelves), and annotated genes (i.e. promoter [TSS], body, UTR, and 1st exon). Distribution of all HM450 probes is shown on the left panels for comparison. **e.** Heatmap of the 71 CpGs associated with *in utero* AFB1 exposure (AfMSs). Unsupervised dendrograms are shown for CpG sites and samples. Annotations in the lower panel illustrate the corresponding aflatoxin exposure and sex. **f.** Methylation correlations for a selection of AfMSs (beta values) and aflatoxin exposure (pg/mg of albumin).

Figure 2. Distribution of inferred cell subpopulations. HM450 array data was used to infer the percentage of each of six different cell subpopulations, as described in Methods. **a.** Inferred data was plotted by aflatoxin exposure category (i.e. Low and High). **b.** Inferred data was also correlated to the absolute value of aflatoxin exposure, based on AF-alb, for each of the six blood cell subpopulations. Significant Pearson correlations ($p<0{\cdot}05$) were observed for CD8T cells and Granulocytes.

Figure S1. Pyrosequencing quantification of DNA methylation is shown for selected CpG sites, comparing Low vs. High aflatoxin *in utero* exposure.