PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Is famine exposure during developmental life in rural Bangladesh associated with a metabolic and epigenetic signature in young adulthood? A historical cohort study.
AUTHORS	Finer, Sarah; Iqbal, Mohd; Lowe, Rob; Ogunkolade, B; Pervin, Sonia; Mathews, Christopher; Smart, Melissa; Rakyan, Vardhman; Alam, Dewan; Hitman, Graham

VERSION 1 - REVIEW

REVIEWER	Elmar W. Tobi Molecular Epidemiology Leiden University Medical Center The Netherlands
REVIEW RETURNED	01-Apr-2016

GENERAL COMMENTS	Sarah Finer et al. present a well-written manuscript on gestational and postnatal famine exposure in Rural Bangladesh. It makes an interesting addition to the field, replicating findings on gestational famine exposure and elevated BMI and replicating findings on periconceptional energy shortages and DNA methylation at meta-stable epialleles. The authors clearly denoted that the sample size is limited in both the Methods and Discussion section, which then results to certain choices in the analysis strategy that offers room for discussion. The reported results on 120min glucose levels, which are given great prominence in the discussion, rely on a subdivision of the data. A greater transparency on the robustness of these findings might help to convince the readers on this point. Moreover, a negative result is presented for an EWAS on these individuals for an analysis that at best can be considered a crude analysis. Replication of findings from other studies using the same 450K platform will hopefully be reconsidered considering the rarity of these types of data and the stressed importance of replication by the authors in Introduction and Discussion. Here I present comments which I hope will strengthen a very interesting and unique study. Major: Power and group based testing. Phenotypes The authors use 3-way anova (and occasionally 2 way) to test for (ordinal) group differences, stating that their limited power excludes multivariate analysis. Small sample sizes reduce the ability to efficiently/correct for possible confounders. However, I do not concur that 191 samples (A+B+C, Table 1) exclude the possibility, nor necessity, to investigate the influence of gender, wealth status etc. on the outcomes presented as part of normal sensitivity analyses.

	Despite not being significantly different between groups, small differences in sex ratio might still influence analyses on BMI as BMI is often heavily influenced by gender. The authors should therefore use (ordinal) regression. Despite explicitly denoting their study as limited in power, the authors then continue with dividing the continues BMI variable into 3 groups (Aiken & West, 1991 argue that categorizing a continues variable, no matter how you do it, is a loss of power). I find the observation on glucose at 120min stratified by these groups intriguing, but I am currently unconvinced (and would like to be convinced). Since group sizes are now down to sometimes 4 individuals the authors should show their (raw) data as to convince the reader that these are not findings driven by individual samples/groups. Figures as Figure 3 in Waterland et al. PLOS genetics 2010 but now for glucose measurements might be good to include so that readers can judge for themselves the robustness of this finding. Power and group based testing. Methylation As it currently stands the presented EWAS is a very crude analysis and in my opinion some additional work is needed to exclude the possibility of any interesting results. Moreover, this should help make the work “beefy” enough to warrant publication. In addition to sample numbers the power here is also determined by the balance between the amounts of nuisance variation in relation to the expected biological variation of interest, which is small for famine/nutrition effects (<5% differences). The first 2-3 principal components in the data are most likely to be cellular heterogeneity, accounting for almost 20% of the total variance. Correctly modelling such variance will most likely boost power, since it costs just <3 df in a multivariate model. Also, despite the randomization across bisulfite batches and arrays some correlation might have arisen between the outcomes and the technical effects by chance. This is more often so than not so when inspecting –omics datasets. Latent variables could be used that summarize multiple of such nuisance effects simultaneously. A mathematically powerful approach is coded in the R package “cate” (and easy to execute). Generally one can generate a few latent variables that correct for cell heterogeneity and between array and on-array technical effects with minimal costs in terms of df added to a multivariate model. Likewise an option is restricting the analyses to the proportion of the data for which a-priori effects are expected. Enhancers, CGI-less promoters and meta-stable epialleles might be extracted from the total datasets for a semi-candidate approach. Meta-stable epialleles I was a bit puzzled, I must confess, by the reporting on the meta-stable epialleles. In the first study all 5 tested MEs were significant, this dropped to one in the 2014 study. A composite z-score was calculated that was significant in both studies. Only RBM46 was significant in the 2014 study (and also found in another, infancy exposure study: http://www.sciencedirect.com/science/article/pii/S0006322316322338). Therefore I feel that the authors should rewrite their discussion on this point to better reflect the overall replication rate across all studies. In this respect I think it is vital to also calculate & include the overall mean z-score to show that overall the results are similar/identical across all 3 studies. (Of note, Dominguez-Salas uses a sex-

	corrected z-score for the mean methylation across MEs [N=120]). This z-score seems to have as an additional advantage that it seems to level the difference in variance between groups apparent for some individual MEs. From the authors' boxplots it is clear that there is no equal variance across groups for several of the MEs in this study, which may influence results. Important Additional clarification of the nature of famine exposure It is currently unclear from the text what the nature and extend of the famine was. The text states income depression as a contributing factor to the famine, suggesting that some food was to be had for the wealthy. How exactly was famine exposure attained from the hospital records? And is there any data on the wealth status of the mother, as this might apparently be a confounder? Which probes are analyzed? Missing are the cgxxxxxx numbers of the probes analyzed in Figure 1. I also missed how Beta values were combined across probes: I assume by mean values? Additional replication? Introduction: "These so-called 'metastable epialleles' occur at regions of the epigenome established stochastically in the early embryo, maintained across different cell lineages, and show intra-individual and inter-tissue stability [19,22] …, we chose to investigate these metastable epialleles due to the robust nature in which they were first identified and characterisation in a contrasting ethnic population, meaning that replication in our cohort would be of potential significance." Following this reasoning arguably the most robust meta-stable epiallele is missing, namely the VTRNA2-1 [Silver et al. Genome Biology 2015]. It was identified using 2 different genome-wide technologies [covering cg11608150 till cg08745965 on the 450K array] and was validated by pyrosequencing across 3 CpGs. In the Gambia the exposure is a combination of low food availability and requirement of hard work on the land, resulting in a negative energy balance. In the current study famine is studied. Famine was likewise studied by this reviewer using the same platform (Tobi et al. IJE 2015, gestational exposure, N=811) in a likewise contrasting ethnic cohort, namely white Caucasians. Given the fact that famine is the exposure in both studies it would be of interest to see what the results are for at least the two CpGs associated with gestational famine exposure in the Bangladeshi cohort. Reading the discussion I interpret the reluctance to test these few CpGs might come from the following interpretation of the Dutch Famine studies: page 6, line 44 "Other studies include postnatally exposed individuals as 'unexposed' controls….". [my apologies for the reverse rebuttal; but I feel inclined to add some clarification]. Indeed, the Dutch Famine cohorts contain some individuals exposed before <2y to famine (some present in the 'time-controls' in IJE 2015 for instance). Therefore we denote the control individuals as "gestational famine unexposed" or "time controls". In the referred studies [17,18] we were looking for gestational exposure specific

	DMRs, and in particular those specific to early gestation. As part of the sensitivity analyses I also compared the exposed to the time-controls only, finding that DNAm differences became even more pronounced if anything. Also for the IJE results exclusion of infancy exposed individuals did not change the result. This is perhaps not surprising as the AMC and Columbia/Leiden Dutch Famine cohorts were specifically set up to test for gestational specific effects (and gestational timing effects) and contain few infancy exposed individuals. The cohorts were thus set up with a different research question, and I cannot recall that AMC/Columbia|Leiden studies ever pretended to excluded effects in childhood. To the contrary, we have contributed to explorations in other cohorts on infancy effects on DNAm of famine DMRs (Obermann-Borst et al. Pediatr Res. 2013). Indeed, literature suggests significant effects from famine exposure in infancy (of note: consistently smaller effect estimates than gestational exposure) on BMI/glucose metabolism/T2D (great leap forward: Li et al. 2010, Wang et al.2015). Some evidence for this was also found for Dutch Famine exposure in the Prospect-EPIC cohort (van Abeelen et al 2012), but not for national registry data in relation to other outcomes (for instance studies by van Poppel or Lindeboom) and also not for the Holodomor famine and T2D incidence (N=200K, Lumey et al. 2015). I concur with the authors: replication is important and given that DNAm studies on famine are rare I believe a unique opportunity is missed by omitting famine results as reported in Tobi et al. IJE 2015 for a very limited number of probes. In relation to the importance of replication; perhaps also of interest are the results on infancy exposure to malnutrition in Barbados: (http://www.sciencedirect.com/science/article/pii/S0006322316322338) In which they also report that one meta-stable epialleles is associated (RBM46) and a famine DMR is likewise associated in their study: GNGT2/ABI3 promoter, hg18, chr17:44,641,718-44,642,919 Minor: Methods: 1. Several variables are discussed that are not analyzed in the paper; for instance T2D. 2. Some summary on how the 450K arrays were pre-processed might be in order; despite referring to [ref 28] some info might be handy for the reader. 3. I assume that baseline and 120min glucose values were log-transformed in the analyses, as is standardly applied in the field because they are normally skewed? No mention is made in the Methods. 4. Data submission: Will the data be deposited in a repository? typos etc. 5. Dominguez-salas et al. 2014 should replace the [19] reference, currently a 2013 paper not on MEs is mentioned. 6. Might want to consider a simple legend in figure 1, rather than only in describing the colours in the text, as it helps in quickly interpreting the graph 7. Bisulfite treatment: line 33 and line 47. 8. line 33: and assessed using……? 9. Line 55/56: crop failure == poor harvest?

REVIEWER	Susanne de Rooij AMC, the Netherlands
REVIEW RETURNED	06-Apr-2016

GENERAL COMMENTS	The present manuscript describes the results of a study on the consequences of a Bangladeshi famine during early life on obesity/diabetes in adulthood and epigenetic differences. The authors conclude that famine during gestation programs towards diabetes in later life and that famine in infancy leads to obesity and diabetes. Epigenetic differences were also found. Although I find this an interesting manuscript with interesting findings, especially the combination of being underweight and having increased diabetes risk after prenatal famine exposure are intriguing, I think that there are a number of major and minor things that have to be improved on. Below, my main points. Major: 1. The study needs better argumentation for the choice of candidate loci to be looked at in the epigenetic study. Why were only the loci from the Gambia study taken into account and not those from the Leiden study performed by Tobi and Heijmans? 2. Much more information on the Bangladeshi famine is needed; there is almost none now. How long did the famine take? How much did people have for food? How was it for pregnant women and small children? Was everybody equally affected? 3. Also more information on the definition of famine exposure is needed. Wat were inclusion criteria? Date of birth and also residency? In the prenatally exposed group, how long before conception could you be exposed? 4. It needs to be stressed more that it is a limitation that especially the prenatal groupw was very small and that no difference could be made between before conception, around conception, early, middle and late gestation. 5. What was the argumentation for doing epigenetic analyses in a subsample of a study group that was already not very large to begin with? 6. A table with descriptive data is missing (information per group on sex, SES smoking etc). 7. I think it has to be mentioned in the abstract that prenatal famine exposure led to underweight and higher diabetes risk at the same time. This is an interesting finding. 8. In the Results section, I do not really understand the sentence ‘data from this background population also highlights the high burden of hyperglycemic disorders’. Is this meant to be saying that in general there is a high prevalence apart from any other study findings? 9. The epigenetic analyses puzzle me. Why were groups A and B combined? 10. Why were there so little people left in the targeted approach? 30, 13 and 18 while in the genome wide group there were 49, 40 and 54? 11. In the discussion, line 24, the authors speak about clear definition of exposure windows. I do not really agree with this. Windows were not so clear. 12. Also in the discussion, line 33, it is said that this appears to be

	independent of confounding factors. Have the authors tested this then? Where are the results shown? 13. The postnatal findings is not in line with evidence form the Dutch famine studies, which show no increased health risk of postnatal famine exposure. In contrast to what the authors say, this has been taken into account, and it does not lead to 'misinterpretation'. 14. In the Discussion, from line 53 on, a number of limitation of other epigenetic studies are summed up, however, I do not see how the present study, although interesting, overcomes the most of these, apart from the replication and validation. Surrogate tissues are also used and the causal or clinical phenotype effect is also not really tested in this study, in the sense that epigenetic differences are not related to the phenotypes (why not?). 15. The conclusion at the end of the Discussion (line 32) does not seem correct to me. The postnatally exposed were over and not underweight? 16. Line 42, 43 in the discussion: I do not think the authors have taken ethnicity and SES into account? Minor: 1. It sounds a bit weird speaking of developmental life, what about changing it to early life? (for example in title). 2. A good look has to be taken at grammar and spelling. 3. In conclusions of the abstract, it is said that 'maternal famine exposure..', however in case of the postnatal group the exposure was not maternal. 4. In Strengths and Limitations, what is a 'stable' population? 5. In Tables 2 and 3 can you please provide the number of participants per group?

VERSION 1 – AUTHOR RESPONSE

Dear Reviewers,

We would like to thank both of you for your helpful comments and the constructive nature in which these have been offered during the review of our manuscript. The issues raised have been addressed and we have followed all guidance and suggestions where possible and we feel that our manuscript is stronger as a result. We look forward to hearing further feedback on this.

I would like to make the following general points regarding the revisions to our manuscript, before discussing each point from the reviewers in turn.

General points

1. Reviewer 2 has rightly criticised our over-reliance on famine onset and offset dates in our definition of those individuals exposed during gestation as also being 'periconceptually exposed'. We agree that to define a precise window of periconceptual famine exposure in this context is difficult, and we have therefore removed this term from our discussion. In reflection of this, we have made a decision to include some additional clinical data from our gestationally exposed group (group B) that we had excluded in the earlier version of this manuscript due to a gestational exposure of 7 months, rather than a full 9 months. This added data has not changed our outcomes or our methylation data (which has been unaffected by this change), but does give us some extra power in the gestational exposed group from a larger sample size (n=68), and we hope the reviewers agree with our having made this decision
2. We have expanded the range of metastable epialleles studied to include those in the most recent

Gambian papers and are delighted to find that we also replicate some of these more recent findings. We have changed the presentation of our methylation data, as suggested by reviewer 1, to reflect the methylation across each complete metastable epiallele rather than select single probes or groups of probe by genomic feature as we had done in the previous version of the manuscript. We feel that this strengthens our analysis and the suggestion of reviewer 1 to include a z-score mean across all metastable epialleles add further weight to our replication of the Gambian studies. The reviewers will note some subtle differences in the list of Dominguez-Silver MEs that 'replicate' between the first and second version but highlight that this is because we are no longer presenting single probe methylation differences as we did previously. If anything, we feel that our robust re-analysis will be a conservative underestimate of the number of metastable epialleles that are replicable.
3. With their consent, and if accepted, we would like to reference both reviewers by name in our acknowledgements in view of their helpful contribution to our work.

The rest of the reviewers' comments are addressed below, with our responses marked with an asterix (*).

With many thanks,
Yours sincerely,

Sarah Finer

*RESPONSE TO REVIEWER 1
Reviewer 1 comments and authors' response

Sarah Finer et al. present a well-written manuscript on gestational and postnatal famine exposure in Rural Bangladesh. It makes an interesting addition to the field, replicating findings on gestational famine exposure and elevated BMI and replicating findings on periconceptional energy shortages and DNA methylation at meta-stable epialleles. The authors clearly denoted that the sample size is limited in both the Methods and Discussion section, which then results to certain choices in the analysis strategy that offers room for discussion. The reported results on 120min glucose levels, which are given great prominence in the discussion, rely on a subdivision of the data. A greater transparency on the robustness of these findings might help to convince the readers on this point. Moreover, a negative result is presented for an EWAS on these individuals for an analysis that at best can be considered a crude analysis. Replication of findings from other studies using the same 450K platform will hopefully be reconsidered considering the rarity of these types of data and the stressed importance of replication by the authors in Introduction and Discussion. Here I present comments which I hope will strengthen a very interesting and unique study.

Major:

Power and group based testing. Phenotypes
The authors use 3-way anova (and occasionally 2 way) to test for (ordinal) group differences, stating that their limited power excludes multivariate analysis. Small sample sizes reduce the ability to efficiently/correct for possible confounders. However, I do not concur that 191 samples (A+B+C, Table 1) exclude the possibility, nor necessity, to investigate the influence of gender, wealth status etc. on the outcomes presented as part of normal sensitivity analyses. Despite not being significantly

different between groups, small differences in sex ratio might still influence analyses on BMI as BMI is often heavily influenced by gender. The authors should therefore use (ordinal) regression.

1. Ordinal regression of confounders on phenotype – but whether gender, wealth status etc impact could be part of causative pathway.

*RESPONSE: We thank Dr Tobi for this helpful suggestion and have included an ordinal regression analysis in our results, showing that there is no significant effect of these confounders on the BMI phenotype, as follows: “Ordinal regression analysis showed that BMI categorisation was not affected by potentially confounding variables (age at follow up, sex, wealth status, marital status, occupation, educational attainment, smoking or arsenic exposure) (pseudo R-Square 0.053, p = 0.34).”

Despite explicitly denoting their study as limited in power, the authors then continue with dividing the continues BMI variable into 3 groups (Aiken & West, 1991 argue that categorizing a continues variable, no matter how you do it, is a loss of power). I find the observation on glucose at 120min stratified by these groups intriguing, but I am currently unconvinced (and would like to be convinced). Since group sizes are now down to sometimes 4 individuals the authors should show their (raw) data as to convince the reader that these are not findings driven by individual samples/groups. Figures as Figure 3 in Waterland et al. PLOS genetics 2010 but now for glucose measurements might be good to include so that readers can judge for themselves the robustness of this finding.

2. Plot glucose data as requested

*RESPONSE: We concur with the reviewer about the intriguing data presented in this analysis and present the 120minutes glucose data in graphical format (please see figure 1), as requested, to assist the reader. We would also like to highlight that we have been careful not to overstate this finding, or its extrapolation, in our results and discussion sections. To this effect, we also now highlight the need for replication in future studies in our discussion “The finding of impaired glucose tolerance in the underweight exposed in gestation will benefit from replication in future studies and may suggest a programmed….”

3. Power and group based testing. Methylation
As it currently stands the presented EWAS is a very crude analysis and in my opinion some additional work is needed to exclude the possibility of any interesting results. Moreover, this should help make the work “beefy” enough to warrant publication.
In addition to sample numbers the power here is also determined by the balance between the amounts of nuisance variation in relation to the expected biological variation of interest, which is small for famine/nutrition effects (<5% differences). The first 2-3 principal components in the data are most likely to be cellular heterogeneity, accounting for almost 20% of the total variance. Correctly modelling such variance will most likely boost power, since it costs just <3 df in a multivariate model. Also, despite the randomization across bisulfite batches and arrays some correlation might have arisen between the outcomes and the technical effects by chance. This is more often so than not so when inspecting –omics datasets.
Latent variables could be used that summarize multiple of such nuisance effects simultaneously. A mathematically powerful approach is coded in the R package “cate” (and easy to execute). Generally one can generate a few latent variables that correct for cell heterogeneity and between array and on-array technical effects with minimal costs in terms of df added to a multivariate model. Likewise an option is restricting the analyses to the proportion of the data for which a-priori effects are expected. Enhancers, CGI-less promoters and meta-stable epialleles might be extracted from the total datasets for a semi-candidate approach.

*RESPONSE: Again, we are grateful to Dr Tobi for his helpful suggestions. We performed the suggested analysis, using the 'cate' R package and a latent variable approach. We plotted the distribution of the t-statistic and this closely followed the expected N(0,1) distribution but there was some evidence of under dispersion of the test statistic. e.g. follows a N(0,0.85). To estimate the number of latent variables to use we used the bi-cross-validation method supplied in the cate package inputing the slide number of each chip as an indicator of potential batch. This resulted in an output of 4 latent variables. This was then inputed into the cate function and the correction was performed. This produced a corrected distribution of t.statistics much closer to N(0,1). The value of the X^2 test for confounding produced a p-value of 0.085 which suggests there was limited confounding in the study. We saw no significant adjusted p-values which pass genome wide correction with the lowest p-value being 0.07553139 and have included this result in our manuscript "Neither analysis, nor application of a latent variable approach, yielded methylation differences that held up to correction for multiple testing using a false discovery rate (adjusted p<0.05) and we therefore did not proceed with further genome-wide analysis." We therefore have had to reiterate this negative result and feel that further genome-wide analysis of this data set would be uninformative and prone to potential false discovery. We hope to use this genome-wide data in future replication studies, perhaps using meta-analysis, to overcome the limitation of our relatively small sample size and lack of power to identify significant differences.

4. Meta-stable epialleles

I was a bit puzzled, I must confess, by the reporting on the meta-stable epialleles. In the first study all 5 tested MEs were significant, this dropped to one in the 2014 study. A composite z-score was calculated that was significant in both studies. Only RBM46 was significant in the 2014 study (and also found in another, infancy exposure study: http://www.sciencedirect.com/science/article/pii/S0006322316322338). Therefore I feel that the authors should rewrite their discussion on this point to better reflect the overall replication rate across all studies. In this respect I think it is vital to also calculate & include the overall mean z-score to show that overall the results are similar/identical across all 3 studies. (Of note, Dominguez-Salas uses a sex-corrected z-score for the mean methylation across MEs [N=120]). This z-score seems to have as an additional advantage that it seems to level the difference in variance between groups apparent for some individual MEs. From the authors' boxplots it is clear that there is no equal variance across groups for several of the MEs in this study, which may influence results.

*RESPONSE: Thank you for these constructive observations. We have addressed these and expanded our targeted analysis to include additional metastable epialleles (see letter/general points above and further discussion below). We have also taken the suggested approach of calculating a mean z-score across all MEs which adds weight to our approach and findings by showing significant methylation differences between the 3 experimental groups. We have also included a summary table of the mean methylation at the MEs studied (supplementary table 1) and described the z-score of mean methylation across all MEs in our results section and with box plot (Figure 2). Of note, we have changed the presentation of our methylation (beta value) data at individual genes to show mean methylation across all probes within the reported MEs, as has been done by others, e.g. Dominguez-Salasz. Dr Tobi correctly states that there is unequal variance between MEs and we have therefore presented single ME data as means (with comparison by ANOVA, although in fact non-parametric analysis produces the same result), and combined ME data as an analysis of mean z-score.

5. Important

Additional clarification of the nature of famine exposure. It is currently unclear from the text what the nature and extend of the famine was. The text states income depression as a contributing factor to

the famine, suggesting that some food was to be had for the wealthy. How exactly was famine exposure attained from the hospital records? And is there any data on the wealth status of the mother, as this might apparently be a confounder?

*RESPONSE: We agree that we have included insufficient detail regarding the nature of the famine. Please see the responses to Reviewer 2, revisions in the manuscript and the additional reference (Razzaque, 1990) which address this point. “A severe monsoon in 1974 destroyed the majority of the annual rice crop, and in combination with inadequacies in the food distribution system, unemployment and lower purchasing power, a severe famine ensued [24]. In Matlab, this famine was defined as occurring between July 1974 and June 1975, and its severity implicated deficiencies of multiple macronutrients and micronutrients in the diet of the local population during that time [25]. Increased mortality of the general population in Matlab during and immediately after the famine was noted, as was neonatal and infant mortality [24].” (…) “ Famine exposure windows were determined from birth records and using the famine start and end dates previously described. All individuals included in the survey were grouped according to exposure window resulting in the following; Group A (n=81): postnatal famine exposure (born 1 to 2 years before the start of famine); Group B (n=40): famine exposure in gestation (including at least 7 months of famine exposure during gestation) and, Group C (n=70): unexposed, (conceived 6 months to 2 years after famine).

Which probes are analyzed? Missing are the cgxxxxxx numbers of the probes analyzed in Figure 1. I also missed how Beta values were combined across probes: I assume by mean values?
*RESPONSE: An additional table (Supplementary table 2) has been included with the probe IDs. Clarification on the beta values has been described above.

6. Additional replication?
Introduction: “These so-called ‘metastable epialleles’ occur at regions of the epigenome established stochastically in the early embryo, maintained across different cell lineages, and show intra-individual and inter-tissue stability [19,22] …, we chose to investigate these metastable epialleles due to the robust nature in which they were first identified and characterisation in a contrasting ethnic population, meaning that replication in our cohort would be of potential significance.”

Following this reasoning arguably the most robust meta-stable epiallele is missing, namely the VTRNA2-1 [Silver et al. Genome Biology 2015]. It was identified using 2 different genome-wide technologies [covering cg11608150 till cg08745965 on the 450K array] and was validated by pyrosequencing across 3 CpGs.

*RESPONSE: Thank you. We have now updated our analysis to include all of the published metastable epialleles identified in the Gambian famine-exposure studies (i.e. from Dominguez-Salasz 2014 and Silver 2015 papers). We are pleased to report that this extended analysis shows additional replication, including of VTRNA2-1, of MEs in our dataset.

In the Gambia the exposure is a combination of low food availability and requirement of hard work on the land, resulting in a negative energy balance. In the current study famine is studied. Famine was likewise studied by this reviewer using the same platform (Tobi et al. IJE 2015, gestational exposure, N=811) in a likewise contrasting ethnic cohort, namely white Caucasians. Given the fact that famine is the exposure in both studies it would be of interest to see what the results are for at least the two CpGs associated with gestational famine exposure in the Bangladeshi cohort.

Reading the discussion I interpret the reluctance to test these few CpGs might come from the following interpretation of the Dutch Famine studies: page 6, line 44 “Other studies include postnatally exposed individuals as ‘unexposed’ controls….”. [my apologies for the reverse rebuttal; but I feel

inclined to add some clarification]. Indeed, the Dutch Famine cohorts contain some individuals exposed before <2y to famine (some present in the 'time-controls' in IJE 2015 for instance). Therefore we denote the control individuals as "gestational famine unexposed" or "time controls". In the referred studies [17,18] we were looking for gestational exposure specific DMRs, and in particular those specific to early gestation. As part of the sensitivity analyses I also compared the exposed to the time-controls only, finding that DNAm differences became even more pronounced if anything. Also for the IJE results exclusion of infancy exposed individuals did not change the result. This is perhaps not surprising as the AMC and Columbia/Leiden Dutch Famine cohorts were specifically set up to test for gestational specific effects (and gestational timing effects) and contain few infancy exposed individuals. The cohorts were thus set up with a different research question, and I cannot recall that AMC/Columbia|Leiden studies ever pretended to excluded effects in childhood. To the contrary, we have contributed to explorations in other cohorts on infancy effects on DNAm of famine DMRs (Obermann-Borst et al. Pediatr Res. 2013). Indeed, literature suggests significant effects from famine exposure in infancy (of note: consistently smaller effect estimates than gestational exposure) on BMI/glucose metabolism/T2D (great leap forward: Li et al. 2010, Wang et al.2015). Some evidence for this was also found for Dutch Famine exposure in the Prospect-EPIC cohort (van Abeelen et al 2012), but not for national registry data in relation to other outcomes (for instance studies by van Poppel or Lindeboom) and also not for the Holodomor famine and T2D incidence (N=200K, Lumey et al. 2015).

I concur with the authors: replication is important and given that DNAm studies on famine are rare I believe a unique opportunity is missed by omitting famine results as reported in Tobi et al. IJE 2015 for a very limited number of probes. In relation to the importance of replication; perhaps also of interest are the results on infancy exposure to malnutrition in Barbados: (http://www.sciencedirect.com/science/article/pii/S0006322316322338)
In which they also report that one meta-stable epialleles is associated (RBM46) and a famine DMR is likewise associated in their study: GNGT2/ABI3 promoter, hg18, chr17:44,641,718-44,642,919

*RESPONSE: Thank you for these detailed suggestions, and the helpful explanation of the Dutch Famine experimental groups and "time controls". Unfortunately, we were unable to replicate the methylation differences at the regions identified in the reviewer's own IJE paper, nor the Barbados study. We interpret the fact that we have replicated the Gambian MEs, but not the methylation differences in the Dutch and Barbados studies as being inherent to the strength of the MEs identified. We look forward to future replication studies and meta-analyses that are scaled appropriately to replicate the methylation signals identified through the EWAS approach of the Dutch and Barbados studies.

Minor:

Methods:
1. Several variables are discussed that are not analyzed in the paper; for instance T2D.
*RESPONSE: The overall prevalence of type 2 diabetes was low in groups A,B and C and the comparison to the 'older age group' allows us to say that this is likely to be due probably due to their young age at study. We therefore used standard measures of 'pre-phenotype' such as OGTT parameters and these are routinely used in similar studies of diabetes risk. We have clarified the use of these measures in our methods section by saying "Phenotypic differences relevant to type 2 diabetes and obesity were compared between groups A, B and C were examined using three-way ANOVA tests for continuous variables and Chi-squared tests for categorical data. " and refer the reviewer to Table 1 which includes 'Type 2 diabetes' in the last row.

2. Some summary on how the 450K arrays were pre-processed might be in order; despite referring to

[ref 28] some info might be handy for the reader.
*RESPONSE: We have expanded the methods to address this and included some key references, as follows “The Illumina HumanMethylation 450BeadChip (Illumina, Inc, CA, USA) was used to assay DNA methylation at CpG dinucleotides on a genome-wide scale, incorporating all designable RefSeq genes and enriched for CpG islands and regulatory regions. The experimental methods used are previously described [28] and use bisulphite conversion and Infinium bead chemistry [29] to quantify methylation at CpG dinucleotides. Normalised, filtered and quality-checked quantitative methylation data were computed and beta values (0-100% methylation) were derived from the analysis and data were analysed on a genome-wide scale using F-tests between groups A,B and C using Limma [30] and a Benjamini and Hochberg false discovery rate control [31]”

3. I assume that baseline and 120min glucose values were log-transformed in the analyses, as is standardly applied in the field because they are normally skewed? No mention is made in the Methods.
4. *RESPONSE: These were not log transformed as, after breakdown for BMI category, they were normally distributed.

5. Data submission: Will the data be deposited in a repository?
*RESPONSE: Yes, we will make our methylation data profiles available using the GEO repository and have included a data-sharing statement to this effect at the end of the manuscript.

Typos etc.
6. Dominguez-salas et al. 2014 should replace the [19] reference, currently a 2013 paper not on MEs is mentioned. *Done
7. Might want to consider a simple legend in figure 1, rather than only in describing the colours in the text, as it helps in quickly interpreting the graph.
*RESPONSE: Done

8. Bisulfite treatment: line 33 and line 47.
*RESPONSE Done

9. line 33: and assessed using……?
*RESPONSE: Changed to “efficiency assessed in bioinformatics quality control checks.”

10. Line 55/56: crop failure == poor harvest?
*RESPONSE: This has been routinely described as “crop failure” due to external factors, e.g. monsoon.

*

RESPONSE TO REVIEWER 2

Reviewer: 2

Please state any competing interests or state 'None declared':
None declared

Please leave your comments for the authors below
The present manuscript describes the results of a study on the consequences of a Bangladeshi famine during early life on obesity/diabetes in adulthood and epigenetic differences. The authors conclude that famine during gestation programs towards diabetes in later life and that famine in infancy leads to obesity and diabetes. Epigenetic differences were also found.

Although I find this an interesting manuscript with interesting findings, especially the combination of being underweight and having increased diabetes risk after prenatal famine exposure are intriguing, I think that there are a number of major and minor things that have to be improved on. Below, my main points.

Major:

1. The study needs better argumentation for the choice of candidate loci to be looked at in the epigenetic study. Why were only the loci from the Gambia study taken into account and not those from the Leiden study performed by Tobi and Heijmans?

*RESPONSE: This question has been raised by both reviewers and we are grateful to Drs Tobi and de Rooij for highlighting it. In short, we were unable to replicate methylation differences at the loci identified by Tobi and Heijmans. The ability to replicate methylation differences in the Gambian study but not the Dutch and Barbados studies is likely to be due to the former being defined experimentally as 'metastable epialleles' that are, by nature, more easily detectable than the 'methylation variants' of the latter. Whilst this does not say that one is more or less important, it does highlight why we took the approach to replicate primarily the Gambian MEs as we felt that our sample size would be sufficient for this, but that a good quality replication of the Dutch and Barbados studies would need much larger sample sizes. We hope that the latter will be done in the future by joining several studies together and meta-analysing across them.

2. Much more information on the Bangladeshi famine is needed; there is almost none now. How long did the famine take? How much did people have for food? How was it for pregnant women and small children? Was everybody equally affected?

*RESPONSE: Details regarding famine affecting Bangladesh, and this population, is poorly represented in published academic literature, although we have rectified our omission of the paper by Razzaque (1990) in our references. In addition to this paper, there is a wealth of local data relating to the famine in the region studied (Matlab) by co-authors from ICDDR,B on which this study and experimental design was based. In response to Dr de Rooij's helpful comments, we know the famine to have been very severe, and affected all of the local population due to it being caused by a combination of economic, political and meteorological factors. Overall mortality was 62% higher in the famine period, and 31% higher 1-2 years post-famine, compared to 3-4 years after the famine. The chance of dying as a neonate if conceived during famine was 33% higher than those in a comparable non-famine cohort, and early childhood mortality was also increased. There was variation in the severity of famine, with retrospective analysis suggesting sex- and socio-economic biases, but it is not possible to identify the degree of famine exposure on an individual level, and this is a limitation common to other similar studies. We have included some of this detail in the new version of the

manuscript and in responses to reviewer 1.

3. Also more information on the definition of famine exposure is needed. Wat were inclusion criteria? Date of birth and also residency? In the prenatally exposed group, how long before conception could you be exposed?

*RESPONSE: The timing of the famine has been defined as July 1974-June 1975. Our postnatal exposed group (group A) were born at least months prior to the start of the famine, our gestational exposed group (group B) were exposed to famine for at least 7 months, and our unexposed group (group C) were born at least 1 year after the end date of the famine. We feel that these groups are well enough defined to exclude contamination from the start/end dates of the famine. Date of birth is well-documented by birth attendants in Matlab and we are confident that the retrospective use of this data in the household survey will reflect this. Residency of Matlab during the time of famine was determined by the field workers trained to collect accurate geographical information relating to current, and past, demography as part of the wider household survey in Matlab. Please also see related responses to reviewer 1 and the amendments to the manuscript.

4. It needs to be stressed more that it is a limitation that especially the prenatal group was very small and that no difference could be made between before conception, around conception, early, middle and late gestation.
*RESPONSE: We concur that the sample sizes in our groups are modest, although we note that Reviewer 1 considers the reverse. However, we agree that precise characterization of the pre-, peri-, post- conceptual periods, and the early-, middle- and late-gestational windows is likely to be problematic due to some variation in onset and offset of the famine between households/communities etc. We have therefore removed the term ‘periconceptual’ exposure and use the term ‘gestational’ or ‘in utero’ exposure instead. This has also allowed us to include a slightly larger sample set (n=68, rather than n=40) in the ‘gestational exposed’ group as we have included women exposed to famine for 7-8 months of the famine (by dates), rather than our previous inclusion of 9 months. We have also included the following statement in our discussion “future studies will need to define more precise exposure windows to elucidate whether there are variable effects according to more detailed windows of exposure or the frequency and characteristics of the famine itself. We note that all historical cohort and epidemiological studies of famine conditions have the potential to be limited by variable exposure and susceptibility to the conditions of famine on an individual level and this could alter the molecular read-out of exposure”.

5. What was the argumentation for doing epigenetic analyses in a subsample of a study group that was already not very large to begin with?
*RESPONSE: This was a pragmatic decision based on the fact that not all participants of the clinical study gave consent for epigenetic studies to be performed on their samples, and due to limitations incurred by the cost of the epigenetic studies.

6. A table with descriptive data is missing (information per group on sex, SES smoking etc).
*RESPONSE: These data are now included in Supplementary table 1. We felt that this might be too much information for the main paper (and present some of this summarized in Table 1), but we would be more than happy to go with an editorial decision on this

7. I think it has to be mentioned in the abstract that prenatal famine exposure led to underweight and higher diabetes risk at the same time. This is an interesting finding.
*RESPONSE: We have deliberately kept this point as a discussion point due to the very preliminary nature of this finding and its need for replication (as discussed in the response to Reviewer 1).

8. In the Results section, I do not really understand the sentence ‘data from this background

population also highlights the high burden of hyperglycemic disorders'. Is this meant to be saying that in general there is a high prevalence apart from any other study findings?
*RESPONSE: We have provided the prevalence of an older sample of individuals studied in the same survey for comparison with our analysed experimental groups. Although this has not been included in the analysis (due to the difference in ages), we felt this was useful descriptive data to highlight the 'potential' of the local population to develop type 2 diabetes with advancing age. We have revised this description for clarity to, "Comparison of these data with the 'older' group of individuals (mean age 48 years, exposed to famine after age 16 years) suggests that both findings are disproportionate to the prevalence of underweight and overweight in the background (and older) population. Data from this older population also highlights the burden of hyperglycaemic disorders (type 2 diabetes, impaired glucose tolerance and impaired fasting glycaemia) which have a 31% prevalence in this older group (mean age of 48 years) and presents a useful comparison to the younger individuals in groups A, B and C."

9. The epigenetic analyses puzzle me. Why were groups A and B combined?
*RESPONSE: Groups A+B were not combined. In the genome-wide analysis, we investigated whether a combined "control group" of groups A+C would increase our power (and along the same lines as the control groups used in the Dutch Winter Hunger studies) but this did not and have presented this as we thought this might be a useful negative. We did not pursue any further combined group analysis and would be happy to leave this out for clarity if necessary.

10. Why were there so little people left in the targeted approach? 30, 13 and 18 while in the genome wide group there were 49, 40 and 54?
*RESPONSE: Due to the very limited amount of DNA available from participants and there being insufficient DNA for repeated assays, e.g. 450k array followed by BS-pyrosequencing.

11. In the discussion, line 24, the authors speak about clear definition of exposure windows. I do not really agree with this. Windows were not so clear.
*RESPONSE: We agree with this and have addressed this in our response to (4), and have addressed this in the manuscript with a more detailed definition of the famine dates in the manuscript, and hope this is sufficient to allay the reviewer's concern.

12. Also in the discussion, line 33, it is said that this appears to be independent of confounding factors. Have the authors tested this then? Where are the results shown?
*RESPONSE: Please see the results from our ordinal regression analysis, included in the responses to reviewer 1 which do not show any effect of confounding on BMI.

13. The postnatal findings is not in line with evidence form the Dutch famine studies, which show no increased health risk of postnatal famine exposure. In contrast to what the authors say, this has been taken into account, and it does not lead to 'misinterpretation'.
*RESPONSE: Thank you for this helpful clarification, also provided by reviewer 1. We have changed this sentence accordingly to "[35]. Other studies include postnatally exposed individuals as 'unexposed' controls and our data suggests this could lead to bias in their findings [17, 18]"

14. In the Discussion, from line 53 on, a number of limitation of other epigenetic studies are summed up, however, I do not see how the present study, although interesting, overcomes the most of these, apart from the replication and validation. Surrogate tissues are also used and the causal or clinical phenotype effect is also not really tested in this study, in the sense that epigenetic differences are not related to the phenotypes (why not?).
*RESPONSE: Whilst the reviewer is correct in identifying the limitations of many current epigenetic studies, we would like to highlight the (i) we feel our data contributes an important replication study to complement the mostly un-replicated work that is published in this field, (ii) our use of a 'surrogate'

tissue is appropriate in the context of studying metastable epialleles which have been robustly identified as showing highly conserved and consistent methylation patterns across multiple tissues, e.g. Waterland et al (PLoS Genet, 2010), (iii) the identification of a robust, replicated methylation signature associated with early life famine exposure in whole blood offers a useful translational perspective, e.g. as a biomarker in a readily available clinical sample, and that (iv) that we have discussed these limitations in our manuscript and been careful not to overstate our findings.

15. The conclusion at the end of the Discussion (line 32) does not seem correct to me. The postnatally exposed were over and not underweight?
*RESPONSE: We were unable to identify this area of concern

16. Line 42, 43 in the discussion: I do not think the authors have taken ethnicity and SES into account?
*RESPONSE: We have now done this with our multivariate analysis with all recorded variables that could confound. The individuals studied were from one ethnic subgroup (Bangladeshi South Asians) and therefore ethnicity was not included in this analysis.

Minor:

1. It sounds a bit weird speaking of developmental life, what about changing it to early life? (for example in title).
*RESPONSE: With respect, we disagree. The concept of 'developmental programming' is very well-known, reflects the crucial developmental processes that are going on during conception, gestation and postnatal life, and there are >2000 hits on pubmed using this term. Indeed it is reflected by the popular international society, Developmental Origins of Health and Disease. We would argue that whilst "early life" might reflect similar exposure windows to some, others may interpret this differently and consider that this also represents childhood.

2. A good look has to be taken at grammar and spelling.
*RESPONSE: Thank you. We have done this and would be very happy to make any further changes at proof stages

3. In conclusions of the abstract, it is said that 'maternal famine exposure..', however in case of the postnatal group the exposure was not maternal.
*RESPONSE: Thank you. We have changed this sentence to "Famine exposure in developmental life programmed Bangladeshi offspring towards diabetes and obesity in adulthood but gestational and postnatal windows of exposure had variable effects on phenotype."

4. In Strengths and Limitations, what is a 'stable' population?
*RESPONSE: By 'stable' population, we had meant a population that had undergone low rates of migration in to and outwards from the population studied but we are unable to qualify this statement so have removed it from this section. It now reads "A historical cohort study has allowed detailed analysis of the effects of famine exposure windows during gestation and postnatal life."

5. In Tables 2 and 3 can you please provide the number of participants per group?
*RESPONSE: This has been done.

VERSION 2 – REVIEW

REVIEWER	Susanne de Rooij Academic Medical Centre, University of Amserdam, the Netherlands
REVIEW RETURNED	05-Jul-2016

GENERAL COMMENTS	Response to Author's revision. In general, I think that the authors have done a great job in taking up all the comments and this has greatly improved the paper. I also think these results are of high importance and need to published. However, I have some additional comments in answer to the revision that I think need to be addressed. In answer to my previous point 1 to which the authors replied that they were unable to replicate methylation differences at the loci identified by Tobi and the Barbados study and that this is probably due to the fact that they had too little power to detect the methylation variants: I believe this is probably true, but I do not think it is a reason to not discuss these results in the paper. Obviously, the authors have performed these analyses and I think they have to report the results of this in some way and can then go on and discuss that they did not find results because they did not have enough power. Maybe just put it in the Discussion or something, it would be valuable for readers to know that you did the analyses and nothing came out of it. In answer to previous point 13, I think you have to rephrase the sentence saying 'Other studies include postnatally exposed individuals as 'unexposed' controls and our data suggests this could lead to bias in their findings [17, 18]". Maybe you can say that your study shows that is important to look at postnatal effects as well, but I do not think the studies by Tobi et al are biased. The Dutch famine studies show little evidence that the Dutch famine affected health of the children exposed to the famine in infancy, which is probably due to the fact that children under the age of 1 were relatively protected as they received more rations than other people did. In response to my previous point 14, I agree with all the points raised by the authors, however it was not critique to their study I was raising. I was just raising the point that the authors state in the Discussion that the current epigenetics studies are limited by use of surrogate tissues and the causal or clinical phenotype effect not being tested. That is also true but these limitations are not solved by the present study (which the authors are now implying by starting the discussion with this sentence). I agree that this study is important in replicating data and findings are robust. I hope my point is clear now. My point 15 was that conclusion at the end of the Discussion (line 32) does not seem correct, which concern the authors were not able to identify. I will try to explain myself further, I hope I can make it clear. The conclusion is saying that 'we have found clinical evidence that developmental programming towards hyperglycemia in a cohort of underweight young Bangladeshis exposed to famine during gestation and postnatal life '. My idea was that this is a bit confusing, is if I understand correctly it was found that gestational famine exposure was associated with underweight and hyperglycemia and postnatal exposure was associated with overweight and hyperglycemia.

VERSION 2 – AUTHOR RESPONSE

Response to Dr Rooij's review

"In answer to my previous point 1 to which the authors replied that they were unable to replicate methylation differences at the loci identified by Tobi and the Barbados study and that this is probably

due to the fact that they had too little power to detect the methylation variants: I believe this is probably true, but I do not think it is a reason to not discuss these results in the paper. Obviously, the authors have performed these analyses and I think they have to report the results of this in some way and can then go on and discuss that they did not find results because they did not have enough power. Maybe just put it in the Discussion or something, it would be valuable for readers to know that you did the analyses and nothing came out of it. “

We have now included these additional data and added some additional text (highlighted in yellow) in our introduction, methods, and results sections, as follows:

Introduction, paragraph 2
The aetiology of developmental programming is poorly understood, however, recent epigenetic studies have shown differences in DNA methylation (an epigenetic mark) in association with periconceptual exposure to undernutrition [15] and such processes may underlie the gene-environment interactions in the pathogenesis of complex disease [16] and developmental programming [17, 18]. Most early epigenetic studies applied to developmental programming have been limited by lack of replication and/or the use of a candidate gene approach [17]. More recent studies designed to identify metastable epialleles (MEs) or that use an epigenome-wide association study (EWAS) approach have been informative and identify methylation differences according to early gestation/periconceptual exposures [15, 19–23]. The next important step for these data is to replicate across studies in different populations, to identify whether they have a causal role with known programmed phenotypic outcomes such as diabetes and obesity, and whether they contribute to the high burden of these chronic diseases in South Asian populations.

Methods, paragraph 4
The Illumina HumanMethylation 450BeadChip (Illumina, Inc, CA, USA) was used to assay DNA methylation at CpG dinucleotides on a genome-wide scale, incorporating all designable RefSeq genes and enriched for CpG islands and regulatory regions. The experimental methods used are previously described [32] and use bisulfite conversion and Infinium bead chemistry [33] to quantify methylation at CpG dinucleotides. Normalised, filtered and quality-checked quantitative methylation data were computed and beta values (0-100% methylation) were derived from the analysis and data were analysed on a genome-wide scale using F-tests between groups A,B and C using Limma [34] and a Benjamini and Hochberg false discovery rate control [35]. We interrogated probe beta values at probes within sixteen metastable epialleles (VTRNA2-1, PAX8, PRDM9, HLA-DQB2, PLD6, near ZFP57, AKAP12, ATP5B, LRRC14B, SPG20, near BOLA, RBM46, ZFYVE28, EXD3, PARD6G, ZNF678 and ZFYVE28) ‘metastable epialleles’ that showed methylation variation in Gambian children according to nutritional conditions at conception [15, 19, 20]. We also interrogated our data at sites of methylation variation identified in other genome-wide developmental programming studies using a non-metastable epialleles approach in the Dutch Winter Hunger and Barbados famine studies [22, 36, 23].

Results, paragraph 4
We used our 450k array dataset to assess quantitative DNA methylation as beta values (0-1, unmethylated to fully methylated) at probes overlying the metastable epialleles previously. All other array probes within these genes were interrogated to assess regional patterns of methylation. We performed comparative analysis of mean methylation across samples from group A (n=30), group B (n=13), and group C (n=18) using samples in which bisulfite conversion efficiency >95% had been achieved. Methylation differences were identified at six of the sixteen previously reported metastable epialleles (see figure 2), a result significantly different than would be expected by chance ($p<0.05$). At these metastable epialleles, the predominant methylation difference was driven by group B (gestationally exposed) compared to either unexposed and/or postnatally exposed. Comparison using non-parametric testing and with/without the inclusion of SNP probes did not affect the results.

Furthermore, analysis using z-scores of mean methylation across all sixteen metastable epialleles (to reduce the inter-epiallele variance in methylation), showed significant and relative hypomethylation in group B (mean z-score -0.24), compared to A and C (mean z-scores -0.14 and -0.15, respectively) (3-way ANOVA p-value 0.0003, Bonferroni multiple comparison test significant between groups A vs.B and B vs.C). These findings are consistent with the findings of Dominguez-Salasz et al, and Silver et al [19, 20], who showed hypomethylation in Gambian offspring conceived during the rainy season (i.e. during nutritional depletion) and hypothesise that this is due to reduced methyl donor availability. Interestingly, PAX8 showed relative hypermethylation in group B compared to groups A and C, but this did not affect the pattern of mean hypomethylation across all epialleles. Methylation data and probe coordinates from all sixteen metastable epialleles is presented in Supplementary file 1, tables 2 and 3. There was no significant replication of the methylation differences identified in the recent genome-wide Dutch Winter Hunger or Barbados famine studies (see supplementary file 2).

Discussion, paragraph 3
Despite strong clinical and epidemiological evidence for fetal programming, there is little understanding of its molecular basis. Current epigenetic studies in this context are limited by methodological issues, lack of replication and validation, use of surrogate tissues and inability to prove functional or causal effects on phenotype. We sought to improve on these limitations with a carefully designed study during different exposure windows and by replicating and validating our data across technological platforms. Our genome-wide, discovery-based approach was limited by a small sample size and we were unable to control for false discovery in the context of multiple hypothesis testing and we were therefore unable to produce robust findings with this approach. However, we investigated whether the metastable epialleles identified from rigorous epigenetic studies applied to Gambian children exposed to periconceptual nutritional deficiency [15] were also differentially methylated in Bangladeshi adults. We studied the 16 previously identified metastable epialleles using data from our 450k arrays and identified methylation differences at 6 of them. These metastable epialleles showed relative hypomethylation in association with gestational famine exposure, consistent with the observation of hypomethylation at these regions in periconceptually exposed Gambian children. Interestingly, the metastable epialleles at PAX8 and near BOLA, showed relative hypermethylation in those exposed during gestation, and future studies will need to define more precise exposure windows to elucidate whether there are variable effects according to more detailed windows of exposure or the frequency and characteristics of the famine itself. We note that all historical cohort and epidemiological studies of famine conditions have the potential to be limited by variable exposure and susceptibility to the conditions of famine on an individual level and this could alter the molecular read-out. We validated our 450k methylation data at PAX8 using a bisulfite-pyrosequencing assay and found both platforms to correlate strongly. The design of the 450k array, with probes designed to multiple CpG sites in most genes, allowed us to interrogate regional methylation and we showed spreading patterns of methylation variation around the metastable epialleles. PAX8 is a widely-expressed transcription factor that is known to have multiple roles in thyroid cell differentiation and function via transcriptional regulation of a wide gene network [41]. In providing evidence of methylation variation at 6 metastable epialleles in Matlab offspring, we are confident that these epigenetic marks of exposure are unaffected by age at sampling, ethnicity or geographical setting. These findings lend support to the use of whole blood as a means of detecting functionally-relevant, robustly-identified epigenetic variants, such as metastable epialleles, that might have been set down in early development prior to separation of the 3 germ layers, or that are stable across tissues. Our lack of replication of the top hits identified in the Dutch Hunger Winter and Barbados famine studies may be due to differences in study design, ethnicity, or lack of power, and highlight the challenges of replicating small methylation variants from association-based studies.

“In answer to previous point 13, I think you have to rephrase the sentence saying ‘Other studies include postnatally exposed individuals as ‘unexposed’ controls and our data suggests this could lead

to bias in their findings [17, 18]". Maybe you can say that your study shows that is important to look at postnatal effects as well, but I do not think the studies by Tobi et al are biased. The Dutch famine studies show little evidence that the Dutch famine affected health of the children exposed to the famine in infancy, which is probably due to the fact that children under the age of 1 were relatively protected as they received more rations than other people did."

Thank you for this clarification of the Dutch famine studies. We have amended our discussion of this point, as follows:

Discussion, paragraph 2
We identified a 30-32% prevalence of underweight across postnatal exposed, unexposed and older experimental groups, suggesting a non-age dependent background prevalence. In contrast, we observe an excess prevalence of underweight (49%) in young adults exposed to famine in utero, and this appears to be independent of confounding socio-demographic factors. Those adults exposed to famine during both conception and gestation who remain underweight have a higher mean glucose at 120 minutes post-oral glucose challenge, compared to the postnatally exposed or unexposed. The finding of impaired glucose tolerance in the underweight exposed in gestation will benefit from replication in future studies and may suggest a programmed insulin resistance phenotype previously described by others in similar models [2, 4], and supports the concept of malnutrition-related diabetes and the 'thin-fat phenotype' [12, 38]. We also observed that overweight and obesity were twice as common in young adults who had been exposed to famine during postnatal life (<2 years old), compared to those exposed in utero or unexposed. The varied body weight phenotype associated with the in utero and postnatal exposure windows suggests a variable influence on developing organ systems and behavioural systems, and supports our inclusion of a distinct postnatal exposure group to investigate the effects of postnatal programming on appetite mechanisms [39]. Other studies include postnatally exposed individuals as 'unexposed' controls [17, 18] but our data suggests the importance of considering this exposure window as a separate experimental group with that could be susceptible to different epigenetic and phenotypic variation, compared to in utero exposure. Animal and human studies that focus on postnatal life suggest that 'catch-up growth' in the first 2 years of life following in utero growth restriction is associated with childhood obesity [40]. It is important to note that famine exposure at any age is likely to be followed by a period of rapid growth acceleration due to sudden nutritional repletion as famine conditions were reversed. It is not possible to tell from this, or other studies, whether it is the growth restriction, acceleration, or timing of either, that is the determinant of future metabolic disease programming.

"In response to my previous point 14, I agree with all the points raised by the authors, however it was not critique to their study I was raising. I was just raising the point that the authors state in the Discussion that the current epigenetics studies are limited by use of surrogate tissues and the causal or clinical phenotype effect not being tested. That is also true but these limitations are not solved by the present study (which the authors are now implying by starting the discussion with this sentence). I agree that this study is important in replicating data and findings are robust. I hope my point is clear now. "

Many thanks for the clarification – this is very clear now.

"My point 15 was that conclusion at the end of the Discussion (line 32) does not seem correct, which concern the authors were not able to identify. I will try to explain myself further, I hope I can make it clear. The conclusion is saying that 'we have found clinical evidence that developmental programming towards hyperglycemia in a cohort of underweight young Bangladeshis exposed to famine during gestation and postnatal life '. My idea was that this is a bit confusing, is if I understand correctly it was

found that gestational famine exposure was associated with underweight and hyperglycemia and postnatal exposure was associated with overweight and hyperglycemia. “

Thank you for emphasising this point and highlighting a sentence in our manuscript that was confusing and poorly worded. There was indeed an excess of overweight in the postnatally-exposed individuals, but we were not able to make an association between this exposure, the overweight, and the glucose status (potentially due to lack of power). We did however find an association between underweight and glucose status (hyperglycaemia) in those exposed in utero. We have improved the relevant sentences to convey this, as follows:

Discussion, paragraph 5
In conclusion, we have found clinical evidence of developmental programming in young adult Bangladeshis, with varied effects according to exposure during gestation or postnatal life. Young adults who were exposed to famine during gestation and who remained underweight displayed hyperglycaemia, and those who were exposed in postnatal life were more likely to become obese. We have replicated DNA methylation differences at metastable epialleles described in an independent study in those offspring exposed to famine during gestation. Furthermore, we suggest that postnatal exposure to famine may be an important window of exposure to include in this type of study. Our epigenomic study was underpowered to identify new methylation variants using a hypothesis free approach. Our study has overcome some of the concerns raised by many epigenetic studies, namely lack of reproducibility and external replication. However, larger studies are required to overcome false discovery in epigenomic studies and technical variation between platforms and this dataset provides important proof-of-principle and a source of data for future meta-analysis or replication studies. Our findings also highlight the importance of studying the varied clinical characteristics of people at risk of diabetes, and the contribution of different aetiological factors, such as developmental origins, ethnicity, and socio-demographic factors, is crucial to understanding the complex basis of the disease and its increasing global burden. Future studies should also incorporate longitudinal sampling to examine the progression of phenotypes with age, elucidate the stability of the metastable epialleles and to add insight into whether they have an associative or causative role. The identification of metastable epialleles associated with developmental programming may offer a useful molecular signature of early life influences associated with an increased risk of diabetes. Future studies should investigate whether the findings of developmental programming, via an extreme early life exposure such as famine, can be generalised to a wider population at risk of diabetes, e.g. across generations or via less extreme insults. A combined approach of detailed clinical and molecular studies may help to improve diabetes risk prediction in populations undergoing transition, and could help define windows of susceptibility and/or reversibility with which to focus and tailor intervention.

VERSION 3 – REVIEW

REVIEWER	Susanne R. de Rooij Academic Medical Centre at the University of Amsterdam, the Netherlands
REVIEW RETURNED	23-Aug-2016

GENERAL COMMENTS	No further comments, I am happy with the revisions.