Reviewer’s report

Title: Transcriptome-wide analysis of changes in the fetal placenta associated with prenatal arsenic exposure in the New Hampshire Birth Cohort Study

Version: 0 Date: 12 Aug 2019

Reviewer: Marie Vahter

Reviewer's report:

This manuscript concerns associations of fairly low-level arsenic exposure in pregnant women with differentially expressed genes in their placentas collected at delivery. The manuscript is well-written, particularly the sections about the gene expression analyses. However, more information is needed concerning other parts of the study, e.g. urine and placenta sampling and arsenic analysis and evaluation.

A major weakness of the study is the exposure assessment, which is based on the sum concentration of arsenic metabolites in spot urine samples (one from each woman) collected 3-4 months before delivery and the sampling of placenta biopsies. Even the most elegant and detailed measure of gene expression is of no use, is the exposure measure imprecise or irrelevant. The problem is that most of the inorganic arsenic metabolites have very short half-lives in the body (hours-days), why the sum concentration of arsenic metabolites in a spot urine sample is not a very exact measure of the exposure during pregnancy, and even less so, if there is a critical time period for the changes in placental gene expression. Furthermore, the metabolism of inorganic arsenic by methylation is markedly increased during pregnancy (see e.g. Gardner et al., Reprod Toxicol, 2011), which will affect the toxicity (induction of gene expression in this case), as well as the excretion rate. Therefore, urinary arsenic metabolites may be useful for exposure assessment on a group basis only, unless it can be ascertained that the exposure through drinking water and food is constant over time. In the present study, only 24+24 placentas have been analyzed, why the relevance of the exposure assessment must be thoroughly discussed. I would strongly recommend a separate section on arsenic exposure assessment in Materials and Methods, where the choice of exposure measure and its relevance for gene expression in the placenta at delivery is motivated. Also, it should be discussed (in Methods or in Discussion) to what extent the different exposure groups may represent different dietary intakes, and thus exposure to different forms of arsenic, e.g. DMA through rice, as well as other toxic exposures and nutritional factors. Potential confounding? Did the authors compare the different urinary arsenic metabolites between exposure groups?
It would be a much stronger study, had the exposure assessment (maternal urinary arsenic in gestational weeks 26-28) also been based on the arsenic concentrations in the placentas, which apparently have been measured. Punshon et al. (2019) reported a range of 0.006-18.35 ng/g of arsenic in NHBCS placentas. A comparison might have generated additional results and conclusions. Also, such a design would enable an evaluation of other toxic exposures, e.g. cadmium, which is accumulated in the placenta much more than arsenic.

Important, it has to be made clear throughout the manuscript that the results represent associations only, not proven effects.

Title: Is it really the prenatal exposure that is studied? On page 6, line 7, it is mentioned that only maternal urinary arsenic is measured. At least, it has to be explained how this can be interpreted as the fetal exposure.

Abstract: The introduction is far too long and much information missing in methods and results, especially the exposure assessment and more results presented in the manuscript. In particular, how many women were studied? What did the exposure come from? How was it measured? How were placentas collected and treated? I would recommend reporting differentially expressed genes at FDR < 0.05, not p<0.05 in the abstract. The marked difference is not discussed anyhow. How many differentially expressed genes were common for female and male placentas? Were they generally up-regulated or down-regulated? Which genes showed strongest association? Were potential confounders considered?

Background

Page 3, line 20: It is a misunderstanding that studies in areas with prevalent high levels of contamination, e.g. Mexico and Bangladesh, cover high-level arsenic exposure only. In fact, most studies have a wide range of exposure, from very low to very high.

Page 4, first sentence. It is not clear from the writing that 10 µg/L refers to the drinking water guideline.

Page 4, line 18: How is "low birth weight" defined?

Materials and methods

Page 6, line 4: What was the reason for selecting 24+24 placentas? Was any power calculation performed?

Page 6, line 5: Check "each consisting of", which now refers to participants?

Page 6, lines 9-10: As written, I think the exclusions can be questioned. At least, they are not well justified. For example, what is a "urinary arsenic sampling error"? Also, it should be clarified that it was "an outlier" in the gene expression, not in the arsenic exposure. As mentioned above, a section on exposure assessment would markedly facilitate the reading.
Page 6, line 19: How does diethyldithiocarbamate stabilize arsenic species? How would they change at -80°C?

Page 7, lines 2-4: How many samples had measurements below zero? What is meant by "flagged"?

Page 7, line 4: I would recommend not to use "total urinary arsenic" for the sum of inorganic arsenic metabolites. Even if defined here, it might be misleading for those reading other parts of the manuscript. The abbreviations of the different metabolites are already explained on the previous page.

Page 7, lines 10-12: The collection of placentas and placenta biopsies need to be described in much more detail. As the expression "fetal placenta" is frequently used throughout the manuscript, it is assumed that only the fetal part of the placentas is collected. Please, clarify. Who collected the biopsies "at the time of delivery", and how was it done? Biopsy size? Time periods between placenta delivery, biopsy collection, immersion in RNAlater, and freezing?

Page 7, line 19: What "experiments"?

Page 8, lines 4-6: Incomplete sentence. Median of what? Also the following sentence is hard to understand. What is "raw read counts ~ groups"? Was maternal age the only covariate considered (several are mentioned at the bottom of page 5)? How were other potential confounders evaluated? Also, a reference to Figure 1 already here would probably help the reader to follow the analytical approach.

Page 8, line 11-15: Again, how were the potential confounders in the analysis of associations between gene expression and birth weight evaluated and selected?

Page 8, line 18: What effect size? Gene expression or birth weight? What does a 2.1-fold change mean?

Results

Page 10, lines 11-12: The authors need to be more cautious about mentioning "effects" of arsenic. Only associations are evaluated. Also, why is FDR <0.1 used here, but <0.05 elsewhere? Does this mean that there were no gene sets that were significant at <0.05?

Page 12, lines 3-4. Again, how is "low birth weight" defined? Are there really "many" previous studies indicating "low birth weight" in relation to arsenic exposure? No references are given here. "Low birth weight" is also mentioned on lines 10-11, without definition.

Discussion

The whole first paragraph is repetition of previous statements and can be omitted.
Page 14, line 12: The type of studies where "arsenic has been shown to activate the UPR" should be mentioned, especially as the authors conclude that it "has been shown".

I don't have much to comment on the rest of the discussion on the associations with gene expression, which is rather detailed and somewhat uncritical. However, the relevance of the exposure measure needs a thorough discussion, including the stability of the exposure over time and the changes in arsenic metabolism during pregnancy. Also, the time point of induction of gene expression changes and their stability over time need comments. Other potential confounding needs to be considered in the section on limitations (see my comments above).

**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal