Reviewer's report

Title: Altered DNA methylation in liver and subcutaneous and visceral adipose tissues derived from individuals with obesity and type 2 diabetes

Version: 0 Date: 09 Oct 2017

Reviewer: Miles C. Benton

Reviewer's report:

This manuscript by Barajas-Olmos et al., details the exploration and analysis of both epigenetic and gene expression changes between diabetic and non-diabetic obese patients. This was performed across 4 different tissue types (whole blood, subcutaneous and visceral adipose and liver), all tissues are known to have a direct role in metabolic syndrome and obesity. The authors present several candidate genes and pathways identified as being associated with diabetes and the metabolic syndrome, either directly or indirectly via co-morbidity and etiology. The data set while small is very valuable and by nature (hard to obtain human tissues) is of great interest to the scientific community.

I believe the basis for this piece of research is sound and that the authors don't overstate their findings, however it would benefit from additional clarification, analysis and interpretation. I include my comments below.

Major comments:

Page 4 (Background): This section is extremely short and doesn't really define a lot of the key concepts being explored in the manuscript. Ideally there should be a section on methylation and expression, as well as some introduction to the different tissues being explored, their origins and roles in disease.

It started as a minor comment however I consistently found examples/statements which clearly should be referenced throughout the manuscript, some examples (note there are others that need addressing):

Page 4 Line 43: Recent evidence suggests that tissue-specific transcriptional gene regulation by epigenetic factors plays a role in the development of insulin resistance and diabetes among individuals with obesity.

Page 4 Line 53: DNA methylation could provide a link between environmental influences, genetic factors and the development of defects in insulin signaling.
Emerging evidence supports the role of epigenetic mechanisms as a crucial interface between genetic and environmental influences.

There is no mention of adjustment/correction for multiple testing bias in the methods/results. The methods section (page 7/8) mentions p-values (p<0.05) which I assume to be nominal and unadjusted. Understandably the sample size is small therefore power is going to be drastically reduced, but there needs to be mention of an attempt at adjusting for all the multiple testing that is being performed. Likewise for the pathways analysis, p<0.05 is mentioned - these are usually adjusted (depending on the tool) but there is no mention in text of this. There should be a discussions paragraph which details some of the study limitations such as sample size and cell heterozygosity (see later comment).

Page 8 (Results), samples: as this paragraph stands it doesn't really offer much in the way of information. I didn't realise until I had read through the manuscript that there is in fact a table of this phenotypic data (Table 1). This table needs to be referenced/refereed to in this section of the text.

There also appears to be a discrepancy between the numbers of samples. This section mentions 46 (NDO=23 and DO=23), however in the methods this number is never defined. Reading through I'm assuming that the n=23 in each group is a combination of samples across tissue collections - this needs to be made clear in text. It would also be good to highlight how many/which samples have overlapping data. It seems that there are n=16 which would have data for all tissues examined (limited by those samples with available liver), but again this isn't stated in text. If this is the case the abstract should be updated to reflect this, as it is currently misleading by stating that there are 46 individuals which have had liver, both adipose and blood obtained for them.

I like the inclusion of results around intercorrelation of the methylation profiles between tissue types. I do have a few comments on this:

- I am a little sceptical of the high correlation coefficients between these comparisons between tissues. From the scatter plots contained in Figure S2 and my experience running correlations between individuals and tissues, I find it very unlikely that the high correlation statistics are correct. I suggest the authors double check these calculations, as well as make the data publicly available (see my later comment in regards to this) so this can be independently confirmed.

- In line with my above comment I also think it is quite strange that the authors first establish that the tissues are distinctly different (Figure S1), but then try to establish that they are indeed very similar (Figure S2). I'm not completely sure what they are trying to convey here.
I am surprised that the authors were unable to differentiate between SAT and VAT at the complete methylome level. My colleagues and I published a paper recently (https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-017-0344-4) in which we were able to clearly differentiate SAT from VAT based on a single CpG site. This was in 450k data but I would still expect to see tissue differentiation using the 27k panel of CpG sites.

I am surprised not to see any analysis for Differentially Methylation Regions (DMR), both within and between tissues and DO/NDO. DMR analysis has become fairly standard for methylation array based experiments and there are numerous R packages which implement this (minfi/champ, watermelon, RnBeads, …). I would suggest the authors perform an analysis for DMR, it should provide a little more stringency in terms of association statistics (via grouping CpG sites with DMR) and might add additional support for their existing data/results.

There is no mention of heterozygosity of cells within each tissue type. It is very well established that cell-mixtures within tissues provides a source of experimental 'noise' when performing these types of analyses. There are established methods for correcting in whole blood for methylation (see the R package minfi and the Houseman method), there are also methods available for array based gene expression data. The authors should at least discuss this issue within the manuscript.

There is no mention or attempt to validate results, either by wetlab (real-time, pyrosequencing) or further exploration in publicly available data. I understand that it can be difficult with these types of human tissue to obtain additional samples to perform such validation in, but it is possible to potentially follow up their whole blood findings, and if there are additional tissue samples that weren't included in this experiment they would also provide a great source for this validation experiment. Alternatively there is a lot of publicly available data (both methylation and expression), so it could be possible to seek validation via this route.

Page 15 (Declarations): I was unable to access either data set via the provided ArrayExpress accession numbers, the data might be marked private. This should be checked and all data made available in accordance to BMC journal requirements.

Minor comments:

Page 4

Line 48:- DNA methylation is one of the main epigenetic factors involved in regulation of gene expression.
The 'main' epigenetic factor, I'm not sure if this is exactly what the authors are trying to say. Methylation is one epigenetic mechanism which has potential to modify gene expression, but it is by no means the main factor involved in regulation of gene expression.

Page 6

Line 50:- We only used the arrays that passed internal controls.

I believe this should read: "We only included samples which passed QC of internal control probes."

Line 52:- CpGs in sexual chromosomes were excluded

This should be sex chromosomes, not sexual.

Page 7

Line 4:- Beta is the ratio of the methylated probe intensity and the sum of methylated and unmethylated probe intensities.

I suggest a slight reword to: "The Beta value represents the ratio of the methylated probe intensity and the sum of methylated and unmethylated probe intensities."

Line 20:- "Arrays were processed following the standard protocol."

Assume that this is the standard Illumina protocol, maybe make this explicit.

Line 34:- "Computation and statistical analyses were performed using the R package, version 3.1"

I suggest changing to: "Computation and statistical analyses were performed using R, version 3.1", R uses packages but isn't a package itself.
Line 52: "Genomic DNA obtained from each tissue sample was analyzed using the Illumina Infinium HumanMethylation27 BeadChip."

This is methods not results.

There are times throughout the Discussion where the authors, i.e. page 12 line 12: "Altered DNA methylation was mainly observed in LT and in VAT, which enable clustering by T2D status [22]."

Surely they are referring to their results and should be referencing a figure/table/results from this experiment? If they are referencing a paper to support a discussion point this should be made clearer.

In the discussion some of the terminology starts to get confused, i.e. on page 12 lines 7, 17, 20 the abbreviations OD and OND are used, these should be DO and NDO (as per the rest of the manuscript).

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics
Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

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