Reviewer's report

Title: Distinct DNA methylation patterns of cognitive impairment and Trisomy 21 in Down Syndrome

Version: 1 Date: 25 September 2013

Reviewer: Joanna Holbrook

Reviewer's report:

1. Is the question posed by the authors well defined?
Yes the questions as I understand them are:
A. What DNA methylation differences are common between the 10 DS subjects and 5 age-sex matched controls.
B. What is DNA methylation marks correspond to degree of cognitive impairment in DS subjects
C. Is APP methylation part of either of the above?

2. Are the methods appropriate and well described?
Infinium data processing was through and included removal of low quality, cross-hybridising and polymorphic probes, inter-sample normalisation and colour adjustment and type1/type2 probe correction.

No technical replicates were mentioned in the report and also no assessment of batch effects even though the 15 biological samples described would not fit on one Infinium array (12 samples /array). Batch effects between arrays have been well documented for Infinium data and could even account for the PCA results the authors assign to blood contamination.

Statistical analysis was standard and included FDR correction, absolute methylation difference filter and non-parametric tests.

I did not understand how the effects of PC1 (supposedly describing blood contamination) were subtracted from the dataset. Were the 47,601 variables that most contributed to PC1 removed? After initial QC 468,702 probes remained, after filtering for polymorphisms and cross-hybridistaion, 468,702-32,494= 436,208 probes remain. The transformed dataset has 388,607 probes so were 47,601 probes removed in the transformation? And how were they chosen? By eigenvalues? What was the cut-off? If this is the true approach , it could theoretically overcorrect (as the authors acknowledge in the discussion) but also undercorrect (as the authors acknowledge in the results. A more through description of methodology and discussion of its limitations is needed.

Shouldn’t the correlation of methylation values with BPT, be done in just the DS samples excluding the controls. Cognitive impairment in DS is likely to have a different mechanism (APP related?) than variation in cognitive function in non-DS
indivduals. Also, given the controls all have high BPT scores, they are likely to drive the correlation results, so the level of overlap seen in figure 4a is surprisingly low.

3. Are the data sound?
The data seems of good quality.

The sample sizes are very small and not perfectly gender matched (50 males is cases, 60% in controls).

The effects of the blood contamination though the authors attempted to correct for it, may confound the results, as the authors themselves acknowledge.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

No, Infinium data should be deposited in a public repository such as GEO. I see no reference to that in the manuscript.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

Yes, the discussion is very well balanced.

6. Are limitations of the work clearly stated?

Yes. Very clearly and thoroughly

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

Yes, to my knowledge

8. Do the title and abstract accurately convey what has been found?

Yes

9. Is the writing acceptable?

Yes the writing is of good quality and easily understood.

Major Compulsory Revisions

1. Shouldn’t the correlation of methylation values with BPT, be done in just the DS samples excluding the controls. Cognitive impairment in DS is likely to have a different mechanism (APP related?) than variation in cognitive function in non-DS individuals. Also, given the controls all have high BPT scores, they are likely to drive the correlation results, so the level of overlap seen in figure 4a is surprisingly low.

2. In a related point, I do not agree with statement at the top of page 8 “This is likely reflective of the fact that the lower praxis individuals still cluster separately” The reason there is a division between the low and high BPT score DS individuals in figure 3C is that the probes depicted were selected to covary with BPT.
3. I did not understand how the effects of PC1 (supposedly describing blood contamination) were subtracted from the dataset. Were the 47,601 variables that most contributed to PC1 removed? After initial QC 468,702 probes remained, after filtering for polymorphisms and cross-hybridisation, 468,702-32,494= 436,208 probes remain. The transformed dataset has 388,607 probes so were 47,601 probes removed in the transformation? And how were they chosen? By eigenvalues? What was the cut-off? If this is the true approach, it could theoretically overcorrect (as the authors acknowledge in the discussion) but also inadequately correct (as the authors acknowledge in the results). A more through description of methodology and discussion of its limitations is needed.

4. Please add information about technical replicates (if any) and array-array variation in the dataset. Were the DS samples spread across two arrays?

Minor Essential Revisions
None

Discretionary Revisions
1. I find the treatment of number of CpGs/gene simplistic. As the authors point out, often the CpGs in the same gene are located close to each other but this is not always the case. Rather than simply counting CpGs/gene, CpGs should be clustered into regions and those regions tested for differential methylation. See Jaffe et al (bump hunting).

2. The authors mention in the discussion the possibility that their signature of cognitive impairment is actually a signature of premature ageing. There are multiple studies available looking at DNA methylation changes associated with ageing in peripheral tissues (in Infinium 450K data). Does the cognitive impairment signature overlap with any of them?

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests