Reviewer's report

Title: Revealing the different biological characteristics of porcine mesenchymal stem cells by global DNA methylation and transcriptome integrated analysis

Version: 3
Date: 6 October 2014
Reviewer number: 4

Reviewer's report:

Wu et al describe a study in which they investigate the molecular mechanisms differentiating mesenchymal stem cells (MSC) derived from bone marrow and umbilical cord of Chinese inbred miniature pig-Wuzhishan pig using integrated DNA methylome and transcriptome information.

* Major Compulsory Revisions *

Given there is no replication in the experiment, the title and abstract describing differences between MSCs from two different tissues is ambitious. Where there is pooling of samples, there is no replication and no opportunity to say what is technical variation and what is biological variation between two samples. The authors need to demonstrate that most of the variation is biological and not technical. Without this, comparisons should be qualitative and not quantitative. This caveat needs to be discussed.

The analytical facets of this study are poorly described and show a distinct lack of statistical rigor. There is absolutely no mention in the manuscript or supplementary material how differentially methylated or differentially expressed regions are defined. P-values are presented and there is no description of the test used to derive those values. As the generation and use of p-values without replication is hazardous, significant detail is needed to ascertain the approach used.

In the methods some description of the MeDIP peak calling and the collection of aligned reads into RPKM counts is needed. Were reads summarised around transcripts or gene loci? Was there any normalization?

What is head-subtelomere and tail-subtelomere? This is mentioned throughout the manuscript with no definition.

The discussion of the gene candidates at the end of the results is verbose and could do with further brevity. The speculation in this section should also be moved to the discussion.

* Minor Essential Revisions *

The term ‘modified’ is used throughout. I assume it means ‘CpG methylated’? I found the term confusing, unconventional and unnecessary. It should be replaced with ‘methylated’ or something similar.

L91-96. While the positive results are reported, it is not clear from the text here
what markers were tested, instead the reader has to refer to the methods. Similarly, for readers unaware of stem cell marker details, what is the rationale for using a mixture of identical and different markers across the two cell types? L138, subtelomeres are briefly defined as “7 Mb from each telomere”. It is unclear what is implied here. Is a ‘subtelomere’ region the range between the start and end chromosome coordinates and 7 Mb from the start and end? L158, degree of what? L161, this whole sentence is confusing. What is the p-value of a ‘whole genome peak’? L211. Where is this FDR value coming from? There was no replication and hence no estimate of sample variance so I can only assume this FDR p-value is the chance of calling a peak? What was the peak caller software? How was the p-value generated? Table S7. What is the FDR value in column O? Without further description, one generally assumes FDR to mean the Benjamini–Hochberg procedure, which is an adjusted p-value, by definition which has a range between 0 and 1. The ‘FDR’ column in Table S7 has values far greater than 1. This suggests a mislabeling. Table S8. Where do the p-values in this table come from? * Discretionary Revisions * L198-202. These two sentences verbosely restatements the same observation, which is itself rather obvious. Genome-wide there will obviously be a negative correlation between CGI methylation and expression. Level of interest: An article whose findings are important to those with closely related research interests Quality of written English: Needs some language corrections before being published Statistical review: Yes, and I have assessed the statistics in my report.