Reviewer's report

Title: Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model

Version: 1 Date: 30 July 2014

Reviewer: Pan Tong

Reviewer's report:

The article by Kogelman et al performed an extensive network-based analysis for 36 RNAseq samples related to obesity in pigs. The experiment was well designed using the F2 pig population and the generated data was very useful to the scientific community.

Major Compulsory Revisions:

(1) Line 544~546 stated that genes were selected from lean, intermediate and obese sub datasets. This is a little confusing to me. Which of the dataset was used to construct the network using WGCNA? Does the network differ among different subsets?

(2) Line 549 said 3532 genes were used to build the network; Line 595 said 8745 genes were used to construct the network. Line 619 said 3101 resulting genes remained after filtering. Please justify the difference.

(3) Please describe the rationale for normalization where samples are corrected for gender (line 537~538).

(4) Line 578 “Using BioMart [106] the associated gene names were detected.” Similarly, in line 611, “Associated genes were detected again using BioMart”. Please clarify what gene names are to be detected. My understanding is that BioMart mostly provides gene annotation.

(5) In line 625~627, the authors chose 1104 genes as candidate regulators using GO categories “transcription factor activity” and “signal transducer activity”. Since the top GO identified in Table 1 were “Osteoclast differentiation”, please explain the rationale for this decision.

(6) In line 634~636, the authors calculated the significance of probabilistic score using a t-test comparing with the scores from randomly assigned regulators. A problem with this approach is that with enough sample size, the t-test would be always significant regardless of whether there is a true difference. A more appropriate way is to use the scores from random regulators as the null distribution and compute empirical P value for each of the three identified regulators.

(7) Has the RNAseq data been deposited to public repositories?

Discretionary Revisions
(1) Overall, how well does the gene expression correlate with the traits? Is it better than random? Does a network-based approach improve the correlation with phenotype?

(2) The authors detected three regulator genes, CCR1, MSR1 and SPI1. How do they associate with the traits?

(3) Figure 2 showed a matrix of correlations as well as P values. Are there significant associations after correcting for multiple testing.

**Level of interest:** An article of outstanding merit and interest 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