Reviewer's report

Title: Genetic Network Identifies Novel Pathways Contributing to Atherosclerosis Susceptibility in the Innominate Artery

Version: 2 Date: 29 April 2014

Reviewer: David Aylor

Reviewer's report:

This is a second-pass analysis of mouse eQTL data. The goals are identification of pathways that regulate vascular inflammation and macrophage function, and to drill down on a specific QTL that is linked to both atherosclerotic lesions and liver gene expression profile. The authors relate the genes in the module using new methods, put the data in the context of other published datasets, and compare gene expression in relevant cell culture experiments. Overall, the paper would benefit from clarification throughout, and careful attention to assumptions and interpretations of the methods.

Minor essential

The text stats that # Half of genes assayed (4485) fit in 10 modules. It is difficult to see 10 modules out of Fig S1 or Fig S2. This could be addressed by drawing the module boundaries in these figures. Why is the within-module correlation so skewed (e.g. red in the lower left only)? This reflects distance to the hub?

Likewise, I see only 8 modules in Fig 1a. I am not sure how to help the reader see 10 but it is essential.

There is no mention in the text of Fig 1b. This is a key figure for understanding the results and conclusions. After introducing Cd44 as the most important candidate gene, Cd44 is not mentioned until the fourth page of the results! As such, some reorganization and new text is merited. As I read it, the key results are that the brown module is most correlated with IA lesions and also contains Cd44. Consider moving p10, ¶2 to the methods, making Fig1b part of Fig 2, or making it a separate figure to print after Fig 2 but before Fig 3.

Three modules are linked to IA lesions. Text says brown, pink, and red. Figure 2A shows brown, salmon, and red (with pink falling below the threshold).

Reanalysis of public microarray data finds that a small portion of the brown module (655 genes) corresponds to genes differentially expression between B6 and apoE-/- at 36 weeks. This module is enriched for core macrophage expressed genes. Primary macrophage cells treated with LPS shows changes in both brown module most-connected genes and cd44, but only in the B6 background (not C3H).

The genetic story is a little confusing. The original QTL experiment was in an apoE-/- background, but the validation experiments are not. It would be helpful to
see this addressed in the text. See minor comment about discussion. The eQTL experiment and the module seem a little out of sync. Is it correct to say the Chr 2 target genes = brown module? Or brown module + other modules?

A figure that pulls together the various data sets in (terms of genetics and tissue type) would be more useful than the pages of bar plots (Fig 3-6), some of which could be moved to the supplement.

I don’t understand the approach described on p15, line 7-9, at all: “Using a binomial probability, assuming that alterations in Cd44 expression should alter 50% of the genes on the list by chance.” This assumption seems unsupported, and the binomial test seems odd. Isn’t this a straightforward ANOVA to compare cd44 + and – macrophages and get a list of differentially expressed genes?

Why does the last section (human tissue) make no mention of the cd44 result that jumps out in Figure 7? Does this mean that cd44 is upstream of inflammation response in this mouse system, but not in the human system? This seems like a plausible interpretation. If so, the authors must address it.

Cosmetic/Discretionary

It is unclear from the introduction that the QTL experiment was done previously. The authors helpfully include the relevant details in the methods. However, I suggest you briefly revise the introduction to clarify that this work is a) previously published; b) involved a high-fat diet ; and c) why (B6 x C3H)F2.apoE-/- is significant. The description of the Chr 2 QTL, including its name (if named), significance, 95% confidence interval, number of genes, etc. should be detailed here, as that is background needed to understand the experiment, not new results.

“locus not observed” (p4,line8) is ambiguous in the sense that a locus (unmodified) is just a place in the genome. maybe use “QTL”, “linkage”, or “relationship”?

“precise” is not used quite correctly (p4,line 16), and is unnecessary.

The QTL methods are not necessary since they are previously published, but are appreciated (p5). Since you are including them, this is a good place for the sample size. Notes on the genotype data would make this complete.

There is an organizational challenge here: the authors start with an analysis of eQTL data, take a turn into cell culture experiments and public data sources (mostly to justify focus on the brown module), then return to the original data with the NEO analysis. I think this is how we read the shift on p14, line 21?

Positional candidate cd44 not highly connected within the brown module. However, it is the only gene with a high NEO score, which is a rough likelihood of causal relationship. This is coming from the original eQTL data? The conclusion as I understand it is that Chr 2/cd44 is upstream of the highly connected genes, which in turn are upstream of the bulk of module genes. If this is correct, it should be stated more explicitly.
“confirm novel interactions” (p16). If they are novel, maybe there is a better word than confirm?

The discussion is simply a summary of the results. It would be much improved if the authors described the entire model as they propose it, e.g. a polymorphism near cd44 regulates the expression of cd44, which then regulates… etc.

**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