Reviewer’s report

Title: Characterizing Human Genomic Coevolution in Locus-Gene Regulatory Interactions

Version: 0 Date: 03 Oct 2018

Reviewer: Michael Imbeault

Reviewer’s report:

I have read the paper entitled "Characterizing Human Genomic Coevolution in Locus-Gene Regulatory Interactions" submitted to BioData Mining - I provide here comments, suggestions and criticisms of the work as an expert in genomics.

The introduction is very well-written and gives an appropriate background on coevolution. If anything in the current version it is a bit too detailed, especially at the end where a description of other approaches to coevolution detection (i.e. matrixmatchmaker) would be best suited for a review, as this is not particularly relevant to any ideas expressed in the paper, neither any comparison of approaches is made.

The manuscript describes a coevolution metric for eQTLs, show that a subset of eQTLs have high coevolution scores with their associated gene / promoter and that this could in theory be used to restrict analysis to a subset with an arbitrarily chosen coevolution threshold, improving detection rates by significantly reducing the number of tested SNPs - less tests, less harsh multiple correction, better p-value thresholds, improved detection rates.

In my opinion the coevolution scores for eQTLs provide valuable information to better understand which parts of the vast non-coding genome are susceptible to be biologically functional. Overall the manuscript makes a significant scientific contribution and should be accepted after the following revisions.

I am however not sure if I love the idea of using them to prioritize / restrict analysis to a subset eQTLs with high coevolution scores. Coevolution can mean that both the eQTL and its gene / promoter are highly conserved or are both rapidly evolving - by restricting the analysis to these two extremes you might lose bona fide signals - not everything is either a housekeeping gene, or a recent fast-evolving paralog - for example, you wouldn't want to bias an analysis toward highly conserved genes with highly conserved enhancers in every scenario.

I would have liked the manuscript to focus more on the clear definition / description of coevolving eQTLs in humans.

For example, what is the distribution of conservation scores in relationship with coevolution scores?
What is the distribution of distance between SNP and promoter in studied eQTLs? What is the relationship of that distance and coevolution scores? This one is critical, as no pairs of SNP and promoter could be coevolving because the distance between them is very short (linkage).

Also, show the distribution of SNP allele frequency and coevolution scores (using binned coevolution scores violinplots for example).

Provide violinplot of coevolution bins and eQTL p-values (corrected / uncorrected).

What is the status of repeated elements in the eQTL database you used (repeatmasked or not?).

I would have enjoyed a small section on computational complexity / computing time / memory requirements, even in the material and methods section.

A discussion of other conservation metrics (GERP, Phastcons) compared to the chosen one (PhyloP) to justify this decision.

Moreover, the current figures are not publication quality - indeed they are very far from it and look like the default output from Excel, which wouldn't even be suitable in lab meetings. Please rework all figures, including improvements such as shaded standard deviations for smoothed profiles, as well as basic figure elements such as scale units on axis, titles, etc.

The SVD figure is not particularly informative and if a limit of allowed figures is reached I would shift that one to supplemental at the benefit of others. Most figures in the current paper could easily be merged in one figure with multiple panels.

Finally, a minor edit to the text: Page 3 - line 9 - "identifying all of the functional" - remove "all"

**Level of interest**
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**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

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

No