

# Improving transcriptome assembly through error correction of high-throughput sequence reads

Matthew D MacManes<sup>1\*</sup> and Michael B. Eisen<sup>1,2</sup>

**1** UC Berkeley. California Institute of Quantitative Biology, Berkeley, CA, USA

**2** Howard Hughes Medical Institute

\* Corresponding author: [macmanes@gmail.com](mailto:macmanes@gmail.com), Twitter: [@PeroMHC](https://twitter.com/PeroMHC)

## Response to Editor Comments

Please find the attached paper, which is a revision of manuscript number 2013:04:462:0:0. The manuscript has undergone the recommended changes, except what has been noted below.

Best, Matt MacManes

### 1. About centralizing data and resources.

I have done a couple of things to speak to this issue. Figshare will not accept large files, so I cannot put the read data there. Prior to becoming aware of that limitation, I uploaded the assembly data. Those have a stable, permanent doi, and I think are OK to be left there. Apparently, identical resources should not have different doi's, so uploading them to Dryad is apparently improper.

The read data are uploaded to Dryad <http://dx.doi.org/10.5061/dryad.km540>, thought the link will not become live until paper is published.

To make traversing a couple of difference resources easier for readers, I have made a README which contains all the various links and locations. This readme is located (1) on github and (2) as supplementary material. I prefer the github resource, as scripts can be forked by users, modified, improved, etc. It is versioned, so record of the initial state is maintained.

I moved all the 'gist' scripts to a single folder on github, thus centralizing all those resources. They are also available as supplementary material, though again, I prefer github.

The empirical data was NOT uploaded to Dryad.. Those data are copyrighted by the Broad, and the Dryad CC0 is a more permissive license, so I don't think that uploading is legit. Instead, what I have done is provided a script that allows users to recreate the subsamples dataset from the full dataset that I link to. This script is available on github and as supplementary.

### 2. The dashes have been removed.

### 3. There are 2 supplementary figures, or at least I hope there are.. I did upload as such.

I am happy to make them part of the main document, but I figured that 7 figures for a short MS is too much, and these are important only in that they characterize the simulated datasets, but do speak to findings.

4. I added some text about what made Reptile laborious. In the not included README file, some instructions about how to select parameters if included. Still the best resource for this is the Reptile paper and download.
5. I added some discussion about Echo, that it took way too long to be included in the paper. I considered some speculation about the other correctors, as you suggested. I really don't have any evidence to support my claims, so I elected to leave those ideas out.