Reviewer Report

Title: High precision registration between zebrafish brain atlases using symmetric diffeomorphic normalization

Version: Original Submission **Date:** 4/19/2017

Reviewer name: Carlos Castro-Gonzalez

Reviewer Comments to Author:

The authors introduce a new method to co-register ZBB and Z-Brain, two existing 6 days post fertilization (dpf) larval zebrafish brain atlases, by using the diffeomorphic algorithm SyN in the ANTs software package. With this method, they provide a quick way to aggregate information from different sources into the same spatial framework creating a comprehensive digital atlas that would provide researchers with a unified resource to gain deeper insights from correlations between neural cell identity, connectivity, gene expression and function within the brain.

Digital brain atlases have been generally restricted to the data produced by the research groups that generated them. In this sense, this work is one of the first efforts to merge two of the main databases for vertebrate larval brain development (6dpf zebrafish brain). From this perspective, the work is novel and of enormous utility to the research community.

The manuscript is sometimes difficult to read and the methods and data are not always properly described. I motivate these points below:

* Through the text it is a bit difficult to follow which datasets (and how many) are used to evaluate registration accuracy and which ones were finally included in the atlas for each of the presented results (registration of live scans, ZBB1.2, registration of fixed scans and Z-Brain/ZBB inter-atlas registration). It could be useful to the reader to expand the "Materials & Methods" and Table 3 to summarize, for each of the mentioned results, which datasets were used as templates, how many brain scans were used for evaluation and how many were mapped to the final resource including the gene patterns they contained, which patterns guided the registration in each case and how many repetitions of each are finally available in the database.
* From this perspective, it could also help to include a Figure showing, face to face, the average brain representations of all the ZBB patterns that were analyzed in the paper vs. their corresponding Z-Brain counterparts (similarly to Fig.2 in [Marquart et al. 2015]) and a few examples of how they looked before/after registration.
* A Figure summarizing the final workflow employed could also help the reader: which channels/patterns are finally used to guide the registration, how the rigid and affine steps are used before the elastic transformation, how the alignment is evaluated, etc. (similarly to Fig.3 in [Ronneberger et al. 2012]).
* The "Methods" section describes the elastic registrations performed with ANTs and CMTK but it

doesn't mention how the initial rigid and affine transformations are performed.
* The text mentions a "computational measure" on 12 identified landmarks and a "manual measure" on 10 identified landmarks. However, no details are provided in "Methods" about how those landmarks were identified (automatically? which criteria was given to the 3 blinded reaters?).

In my opinion, some of the evaluation methods employed are not completely aligned to the aims of the study:

* It is a bit confusing why the authors choose to measure (a) mean cross-correlation (MCC) on 50um cubes around 12 landmarks and (b) mean distance to 12 different landmarks. How are the 12 landmarks different to the prior ones? Was it an ad-hoc choice or what was the rationale behind choosing them? Couldn't the same landmarks be used for both measurements? Why are "8 landmarks" and "5-18 landmarks" used for measurement later in the text?
* The text mentions that "parameters which yielded the greatest increase in MCC often produce abnormally elongated cells" which seems a strong indication against using MCC as the reference metric for the method. Indeed, [Rohlfing 2012] mentions that "measures such as [...] image similarity [...] do not provide valid evidence for accurate registrations and should thus not be reported or accepted as such". Among these measure of image similarity, [Rohlfing 2012] includes the use of image cross correlation (CC). Restricting CC measurement to a few image regions around landmarks does not solve the problem in my opinion. [Rohlfing 2012] concludes "of the criteria tested in our study, only overlap -(measured as Jaccard index)- of sufficiently local labeled ROIs could distinguish reasonable from poor registrations. One reason for this is that smaller, more localized ROIs approximate point landmarks, and their overlap thus approximates point-based registration error". From this perspective, measuring the mean landmark distance (MLD) in a number of landmarks distributed to cover the image makes much more sense to me: (a) This metric has already been successfully used in prior atlas literature (see [Ronneberger et al. 2012], (b) this metric aligns much better with the goal of "aligning neurons within a cell diameter (~10um)" specified in your Introduction and (c) it is supported by the conclusions in [Rohlfing 2012]. My recommendation is to remove MCC as an evaluation metric and replace it by MLD throughout the paper.
* Additionally, the MCC reported in the results vary considerably from as low as 0.1 to as high as 0.9. There is no indication to the reader about what is the minimum acceptable value to consider a registration appropriate. The dispersion of the results for different brains in Fig.3 also raises questions about the robustness of the method when using different brain scans. In my opinion, these results reflect not only the registration accuracy but also the biological variability between different individuals. From this point of view, it continues to make more sense to me to focus evaluation metrics on MLD where the amount of error that cannot be directly attributed to registration inaccuracies was already quantified (the approximate ~5um differences of blinded raters when labeling landmarks in different datasets).
* Metrics in the paper should not only be reported as averages but also with their variability. For instance, MLD variability across the 10 different landmarks should be reported (to evaluate the robustness of the method in different parts of the brain). Similarly, MLD variability across different brain scans should be reported (to evaluate the robustness of the method to different datasets). A minimum

of 6 brain scans were used in the past to assess such variability [Ronneberger et al. 2012].
* The authors mention the impact the elastic registration has on cell deformation. This is indeed a very relevant point and the qualitative observations performed in the manuscript point in the good direction. However, I feel that these observations are restricted to some anecdotal instances and a more quantitative evaluation may be required to back up claims like "cell morphology remained intact". Similarly to selecting 10 landmarks and measuring their MLD, 10 cells -distributed throughout the brain- could be manually segmented before and after registration to quantitatively measure their deformation (e.g. using the Hausdorff distance [Zanella et al. 2010]). The parameter optimization could then be guided by the objective of achieving an MLD<10um while minimizing cell deformation.
* Overall, evaluation criteria (MLD, MCC, visually-observed deformation, M1, M2) and number of datasets evaluated (e.g. 6 brains used in Fig.3 vs. 1 brain used in Fig.1a vs. 3 brains employed in Fig.1e-b vs. 167 in Fig.2, etc.) seem to be really heterogeneous throughout the text. It could help to have a consistent unified criteria for the whole paper (e.g. quantitative evaluation of MLD and cell deformation in, say, 2 sets of 3 larvae every time parameters are optimized).

Some of the conclusions are not adequately supported by the data shown,

* ZBB1.2 (with ANTs) is claimed to have "improved registration precision" compared to ZBB (with CMTK). However, the slight improvement reported is not statistically significant. Under these circumstances, it may make more sense to call these results comparable.
* The claims about cell deformation are based on subjective judgments about registration quality. A more systematic/quantitative approach may be required to generate the supporting evidence.

Regarding the journal's guidelines on minimum standards of reporting,

* The exact sample size (n) for each experimental group/condition is not clearly reported in the Methods section (see comment above). For instance, for the "fixed registration" and "inter-atlas registration" sections, it is unclear which datasets are used (the 167 brains generated by the group vs. the 197 Z-Brain tERK -which in the intro was reported to contain 899 scans).
* Summary statistics alone are reported sometimes (e.g. aggregate average values) without showing individual data values.

Minor comments:

* Caption in Fig.2 explains what the arrow points to in (d,e) but not in (f,g), (h,i) and (j,k).
* In Fig.3, specify which of the data points in (e,f) corresponds to the dataset shown in (g).

References:

[Marquart et al. 2015] "A 3D searchable database of transgenic zebrafish Gal4 and Cre lines for functional neuroanatomy studies." Frontiers in neural circuits 9 (2015).
[Rohlfing 2012] "Image similarity and tissue overlaps as surrogates for image registration accuracy:

widely used but unreliable." IEEE transactions on medical imaging 31.2 (2012): 153-163.
[Ronneberger et al. 2012] "ViBE-Z: a framework for 3D virtual colocalization analysis in zebrafish larval brains." Nature Methods 9.7 (2012): 735-742.
[Zanella et al. 2010] "Cells segmentation from 3-D confocal images of early zebrafish embryogenesis." IEEE transactions on Image Processing 19.3 (2010): 770-781.

Methods

Are the methods appropriate to the aims of the study, are they well described, and are necessary controls included? No

Conclusions

Are the conclusions adequately supported by the data shown? No

Reporting Standards

Does the manuscript adhere to the journal's guidelines on minimum standards of reporting? NoChoose an item.

Statistics

Are you able to assess all statistics in the manuscript, including the appropriateness of statistical tests used? Yes, and I have assessed the statistics in my report.

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:

- 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?
- 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?
- Do you hold or are you currently applying for any patents relating to the content of the manuscript?
- 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?
- Do you have any other financial competing interests?
- 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

To further support our reviewers, we have joined with Publons, where you can gain additional credit to further highlight your hard work (see: https://publons.com/journal/530/gigascience). On publication of this paper, your review will be automatically added to Publons, you can then choose whether or not to claim your Publons credit. I understand this statement.

Yes