Peer Review File

Manuscript Title: Tools for connectomic reconstruction and analysis of a female Drosophila ventral nerve cord

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

This paper beautifully illustrates the way connectome at synaptic level can contribute to decipher complex neuronal circuits/behavior by linking many imaging modalities and literature. Based on the EM volume of the Drosophila Female Adult Nerve Cord (FANC), authors describe the application of new software tools for cell-type annotation, nuclei identification, determining connectivity via synapse identification, and identification of genetic driver lines based on reconstructed neuronal morphology. While annotation is an ongoing process, authors share their tools with the community to tackle different biological questions.
They illustrate the strength of the approach by analyzing at synaptic level the circuits that coordinate the wings with the middle and front legs during escape takeoff.

The paper describes the combination of novel approaches and novel tools to study complex neuronal circuits controlling muscle movements. Based on existing literature having used others methods, they consolidate the know-how in the described circuit adding the number, target and types of synapses. The approach based on multiple techniques as well as many references in literature is solid. The quality of data is as good as it can get right now in the field and the presentation of data is didactic as well as easy to consult via the tables directly directing readers to a Neuroglancer instance showing the determined neuron.
While it is a work in progress, the data in this paper is sufficient to evaluate the strength of the method and to illustrate it with a biological relevant example.
I cannot judge the statistical approaches not being an expert in such analysis.
The conclusions are supported by the data and, most importantly, the paper is opening the access to this dataset for the community by sharing the data as well as the tools necessary to analyze it.
It is without doubt a milestone in the connectome field.

The paper is well written, clear even for non-experts in the field. Experiments and results correspond to the statements in the abstract and conclusion. the introduction gives credit to all the relevant previous work at my knowledge.

Minor revisions:

line 27 : draft - what does it mean at the time of publication (or link to a table updating the advance of the work...)
line 89 ; what are the "false positive" at the ultrastructural level ? are they others cell types ?
line 104 : how do authors define "major errors" ? what types of "minor errors" are not detected ?
line 105-115 and figure 2 : as this paper is about how crucial it is to have the synapses, the part about synapse detection could have more quantitative data instead of "most of" , majority of, etc.. Having some percentages/ quantification of the errors would help to understand the strengths of this automatic/proof-reading approach.
Fig 2A : highlighting the mismatch in one of the 2 panel will make it clearer for reader
fig 2B : quantification of the number of branches missed, their length and/or branching order could make the example more powerful.
fig 2C : having some numbers would help to understand the accuracy as the illustration gives the (wrong) impression that most of tte synapses are mismatched.
Fig 4 and 5 : link to the synapses or some EM images in the figure would illustrate the type of synapses identified.

mat and methods : a description of the synapse annotation/proofreading is missing

Fig S4 : legend for panel E missing

Referee #2 (Remarks to the Author):

The manuscript, “Tools for comprehensive reconstruction and analysis of Drosophila motor circuits” by Azevedo, Lesser, Mark, and Phelps et al., presents a draft connectome of the Drosophila ventral nerve cord (VNC). This is based on an automated segmentation by zetta.ai of the FANC dataset. Additionally, a proofreading platform based on FlyWire is released, and segmentation and platform will enable researchers to fully proofread the FANC dataset in the coming years. Azevedo et al. then utilize these data, along with datasets of different modalities, specifically sparse genetic driver lines and x-ray data, to study which VNC motoneurons innervate which leg and wing muscles.

Major

The manuscript's focus seems to oscillate between the release of a novel dataset resource (‘the yet-to-be-proofread FANC segmentation’, Fig1 and Fig2), and a study about motor neuron muscle innervation

(Fig3, 4, 5, 6). The unveiling of a non-proofread segmentation/connectome of an EM dataset, which was published a while back, doesn't seem groundbreaking unless the connectome's usefulness without proofreading is clarified. This means the novelty and discovery of the studied biological questions bear significant weight.

The authors assembled an impressive consortium of 26 labs for VNC proofreading, underscoring the necessity of this process. However, there's no estimate of how long it will take to proofread the entire VNC, nor is there any statistical data on the consortium's monthly contributions or the time taken to correct major errors for neurons, beyond numbers for very few cells. More detail on this topic is required.

Although the authors heavily emphasize the quality of the automated reconstructions, they conclude that only automated segmentation supplemented with appropriate proofreading, provides a satisfactory level of accuracy for circuit analysis. The level of proofreading required should be dictated by the question at hand. It might be more beneficial to apply "error bars" on the analyses in Fig3, 4, 5, and 6, instead of starting this topic broadly. Was proofreading essential to arrive at these conclusions later, at all?

"Manual" visual comparison appears to be key in matching MN identity across the imaging modalities, which could be a potential weakness. Is there a way to make this process more quantitative? What is the degree of similarity between the next best possible matches?

Minor

The claim that “studying the compact motor circuits of Drosophila could identify circuit mechanisms conserved across limbed animals” seems overly optimistic. The insect VNC appears to be quite diverse and evolutionarily optimized for a particular species (Niven et al., 2008).

The statement that “our approach of using multiple imaging modalities at different spatial scales can also be used to map the inputs and outputs of connectomes across species“ seems overly broad. This might be true for insect species, but I am doubtful that this approach will work for mammals.

The reference to the study on the adult fly brain (Galili et al., 2022) should be updated to include more recent publications, now that the connectome has been released.

I suggest a more detailed discussion on what other questions, beyond motor control, a proofread VNC dataset might be interesting for. This would underscore the dataset's usefulness beyond this study.

Referee #3 (Remarks to the Author):

The present manuscript describes both tools for circuit mapping and analysis and a specific analysis on the coordination of wing and leg motion when initiating flight. The methods include the Popvych et al. toolkit for alignment, Macrina et al. for segmentation, and Buhmann et al. for synapse detection, and Allen Brain Institute's tools for handling large volumes, proofreading with Neuroglancer. All of these tools are state of the art. A consortium of 26 labs is now proofreading neurons in the nerve cord of the fly, an indication of the impact that this work is already having, and which will eventually be harvested in the form of newly reported circuit maps describing the function of the nerve cord.

Outstanding that the team recovered between 60 and 85% of all cable for neurons of interest, compared to manual reconstruction, for the 4 example motorneurons. The number of synapses is within single-digit percent differences (Figure 2).

Figure 2 also states that "we used a synapse threshold of 5 to determine the total population of synaptic partners to compare." But this threshold, even if it reads sensible to me as an insider in the field, isn't justified in the text.

For quality assessment, 4 motorneurons were traced manually and then compared with the automatic segmentation. The differences and errors found are primarily of omission and not proportionally numerous, and the overall overlap is large. This sets a good basis for the work, assuming all other neurons are reconstructed to a similar level of accuracy where 60 to 85% of the cable, and most synapses, are recovered.

The comparison of 200 hours of manual reconstruction of a neuron versus 2 hours of manual proofreading of an automatically segmented neuron reveals just how far the field has gone. That these two neurons differ only in small ways that are irrelevant for the overall circuit architecture offers a reassuring view of the future of connectomics. Now if only these two hours of work could be automated away altogether; but that's beyond the scope of this paper.

Figure 2F and 2G show a curious difference between the amount of precision and recall in the reconstruction of the 4 example motorneurons. For two of them, precision and recall are almost perfect; for the other two, it's fairly poor. Why is that? Surprisingly, the neurons with less synapses (as per Figure 2E) seems to have worse automated reconstruction. What explains this? Was the quality of the manual segmentation poor for these neurons and therefore the discrepancy is higher?

Interesting that the number of motorneurons controlling a leg are in the order of single-digit numbers. This both enables their identification and assignment to GAL4 lines, and renders the system numerically simple enough that the field has a fair chance to understand how it operates. This ought to be highlighted in the Discussion, generating ammunition for study sections to support this work, as this is evidence of experimental and theoretical tractability leading to high bang for the buck, with implications for engineering disciplines seeking to recreate robotic systems, beyond neuroscience.

Polyinnervation – multiple motorneurons innervating the same muscle fiber – is known to occur in

juvenile mammals but not adults; here, it is found in the fly leg, just like it was reported as well for the fly larval body wall muscles before (see Matthias Langraff lab work).

In which other arthropods (line 182) do motorneurons present central output synapses? A citation is missing.

Figure 4: in the legend, 'C' reads like for panel 'B'. and 'B' for panel 'C'.

The description of the Drosophila wing muscle motorneurons alone would have made an amazing paper; that the paper goes well beyond that renders it foundational, an instant classic. The text is accessible and readable, and the explanations on reconstruction methods and accuracy read more than sufficient to ensure the validity of the data as reference for future studies.

How do the motorneurons relate to the thoracic hemilineage-wise thermogenetic assays in Robin Harris et al. 2015 eLife?

In Figure 6C, the volume of motorneurons is plotted against the number of synapses that they receive. Why the volume? Given proximo-distal tapering, and the position of synapses primarily at the distal tips, likely the better correlation is with amount of cable, particularly with the amount of cable with a Strahler number lower than 3. What motivated plotting against volume?

The first sentence of the discussion needs work: that's not the utility of the connectome but its very definition. The best description of the utility of the connectome I've ever read is in Denk et al. 2012 where the authors write the connectome is great to disambiguate between competing hypotheses of neural function.

The first two paragraphs of the discussion can be dispensed with, as they rehash results and not very well.

If you are to discuss connectomes in the context of linking neural dynamics with motor output and posture, don't miss Heckscher et al. 2015 Neuron
https://www.sciencedirect.com/science/article/pii/S0896627315007667

The larval connectome isn't complete yet: only for the whole brain and SEZ, and for select segments and subsystems of the nerve cord. If the last sentence of the discussion refers to comparisons of the motor systems of the larva and the adult, then indeed such comparisons are now possible.

On Methods, line 314: the correct citation for CATMAID's tracing tools is Schneider-Mizell et al. 2016 eLife. The early paper by Saalfeld et al. 2009 introduced the image browsing plus text annotations only.

The Methods seem overly short, despite the appendix on identifying leg motor neurons. In particular, if another team of researchers wanted to reproduce the results, I am not confident they could merely

from the description. The softwares mentioned are complex and take a lot of work to get up and running, and there are far more parameters than the subset mentioned here. Attaching configuration and setup files, together with links to software repositories and IDs of the specific versions used for the various softwares used could go a long way to address this concern.

On the quality checks mentioned in the methods: if these were done using custom scripts and neurons encoded in CVS files or JSON files, why not attach them here. Would enable verification of the claims. The description provided is superficial and does not enable verification.

Author Rebuttals to Initial Comments:

We thank the reviewers for their insightful comments, which have substantially helped to improve the paper and strengthen the communication of our work.

We have significantly revised the manuscript in response to reviewer feedback. Our specific responses to each comment are included below. One major goal of our revisions was to more clearly convey that the paper describes two key advances. The first is the segmentation and proofreading of neurons and glia, and the automatic detection of synapses, in an EM volume of a female fly VNC. The second is the identification of the muscle targets of leg and wing motor neurons, which is necessary for interpreting the output of the fly central nervous system. In the revised manuscript, we clearly delineate these two objectives in both the Introduction and Results, and we have revised the Discussion to put these contributions in broader context. We also significantly expanded the Methods section. We now include more description of the CNN models and technical details about the automated reconstruction. We also include more information about motor neuron identification.

We believe these revisions have significantly improved the paper and we again thank the reviewers for their attentive feedback. Below is a list of significant changes we made to the figures.

Figure 1g: Flipped view so that the VNC is dorsal-side-up, to match figure panels in the rest of the paper.

Figure 2: a-b, Indicated dendritic regions of motor neurons that required proofreading following the initial segmentation, per reviewer suggestion. C, replotted synapses onto a motor neuron to illustrate distance between manual and automatically predicted synapse locations. E, replotted comparison of the number of manual to automatically detected synapses to account for final proofreading state of motor neurons. F, replotted precision and recall of synaptic partners based on automatic segmentation compared to manual annotation (ground-truth), to reflect current proofreading status.

Figure 5e: The neuron labeled “PSn” is now in the supplement as an uncharacterized efferent neuron.

Figure 6c: We re-calculated the values for cell volume based on a new method to correct for mesh issues at the cell body.

Extended Data Figures

New EDF 1a: We provide an inset showing an example EM image of a “false positive” large diameter axon detected as a cell body.

New EDF 3 provides example EM images comparing the automatic synapse prediction and the manually annotated synapses.

New EDF 6 illustrates the necessity of proofreading on top of the automatic segmentation to identify motor neurons.

Point-by-point response (Our responses in blue)

Referee #1 (Remarks to the Author):

This paper beautifully illustrates the way connectome at synaptic level can contribute to decipher complex neuronal circuits/behavior by linking many imaging modalities and literature. Based on the EM volume of the Drosophila Female Adult Nerve Cord (FANC), authors describe the application of new software tools for cell-type annotation, nuclei identification, determining connectivity via synapse identification, and identification of genetic driver lines based on reconstructed neuronal morphology. While annotation is an ongoing process, authors share their tools with the community to tackle different biological questions.

They illustrate the strength of the approach by analyzing at synaptic level the circuits that coordinate the wings with the middle and front legs during escape takeoff.

The paper describes the combination of novel approaches and novel tools to study complex neuronal circuits controlling muscle movements. Based on existing literature having used others methods, they consolidate the know-how in the described circuit adding the number, target and types of synapses. The approach based on multiple techniques as well as many references in literature is solid. The quality of data is as good as it can get right now in the field and the presentation of data is didactic as well as easy to consult via the tables directly directing readers to a Neuroglancer instance showing the determined neuron.

While it is a work in progress, the data in this paper is sufficient to evaluate the strength of the method and to illustrate it with a biological relevant example.

I cannot judge the statistical approaches not being an expert in such analysis.

The conclusions are supported by the data and, most importantly, the paper is opening the access to this dataset for the community by sharing the data as well as the tools necessary to analyze it.

It is without doubt a milestone in the connectome field.

The paper is well written, clear even for non-experts in the field. Experiments and results correspond to the statements in the abstract and conclusion. the introduction gives credit to all the relevant previous work at my knowledge.

We appreciate the reviewer's feedback and comments. We respond to their comments below.

<u>Minor revisions:</u>

line 27 : draft - what does it mean at the time of publication (or link to a table updating the advance of the work...)

The reviewer is referring to a line in the abstract, “We provide the draft VNC connectome and motor neuron atlas, along with tools for programmatic and interactive access, as community resources.” This minor comment prompted an important question that we address below, namely, how complete is the proofreading and what is the advantage of a community structure for proofreading? Still, the word “draft” may confuse readers, so we have revised the abstract to say, “We provide the reconstruction of VNC circuits and motor neuron atlas, along with tools for programmatic and interactive access, as community resources to support experimental and theoretical studies of how the fly nervous system controls behavior.”

We have also added quantification of how many of the somas are associated with objects that have been proofread (Extended Data Figure 1). We emphasize that the proofreading of the FANC dataset is a community effort by including the sentence “Further, the distributed nature of the community-driven FANC effort is a strength, as it allows the connectome to be continuously refined as each circuit is proofread by experts in that area.” in the Discussion section. (line 252)

line 89 ; what are the "false positive" at the ultrastructural level ? are they others cell types ?

We observed that the nucleus prediction algorithm detected some large diameter axons. Large axons appeared particularly prone to segmentation errors across knife marks. Including such large axons in the training set may improve the algorithm in the future. The algorithm also sometimes found broken organelles or other broken membranes. We have rephrased to state “non-nucleus objects” (line 81), and we have included an example image in Extended Data Figure 1A.

line 104 : how do authors define "major errors" ? what types of "minor errors" are not detected ?

We thank the reviewer for their careful reading of the paper. We have changed this phrase to just “errors” in the main text to avoid confusion (line 96). We have also included a more detailed description of proofreading in the Methods. Briefly, we adopted a heuristic of major vs. minor errors. Major errors included missing somas or missing branches leaving the primary neurite (requiring merging objects), or incorrect merges with glia (requiring splitting objects). Minor errors included distal branches that required merging, or small, incorrectly merged objects that required splitting. We show examples of major errors in the new Extended Data Figure 6.

line 105-115 and figure 2 : as this paper is about how crucial it is to have the synapses, the part about synapse detection could have more quantitative data instead of "most of" , majority of, etc.. Having some percentages/ quantification of the errors would help to understand the strengths of this automatic/proof-reading approach.

We now compare the manually annotated synapses and the automatically detected synapses using more precise language. We now state in the text (line 100), “The number of manually and automatically annotated synapses differed by <9% for the largest tibia flexor neuron and the left accessory tibia flexor (Figure 2e). For the other two MNs, automatic synapse detection found ~50% more synapses for the other two MNs due to incomplete manual tracing in synapse-dense regions of the neuropil (Extended Data Figure 3e)”

We also created Extended Data Figure 3 to illustrate this point. We expand on the purpose of Extended Data Figure 3 in response to another review comment below.

Fig 2A : highlighting the mismatch in one of the 2 panel will make it clearer for reader

Thank you for the suggestion. We have edited Figure 2 to highlight a key mismatch in the right accessory tibia flexor.

fig 2B : quantification of the number of branches missed, their length and/or branching order could make the example more powerful.

This is an excellent suggestion. Colleagues are working on tools to address these types of questions, but currently the skeletonization tools are not reliable enough. For example, they sometimes count segmentation errors, such as knifemarks, as branch points. For now, we will continue to report the fraction of cable recovered by automatic segmentation (Figure 2D) while also showing the two reconstructions (Figure 2B) to convey that the primary branches are fully recovered.

fig 2C : having some numbers would help to understand the accuracy as the illustration gives the (wrong) impression that most of the synapses are mismatched.

Thank you for the suggestion. While updating the figure, we found that the synapses are actually closer together than they originally appeared, due to a bug in our code. We now state in the text (line 103), “The

mean distance between from a manually annotated synapse to a predicted synapse was less than 1.5 um (Extended Data Figure 3b)". We also created Extended Data Figure 3 to provide more quantitative detail. In Extended Data Figure 3c and d, we show that the distributions of manual and predicted synapses along the cardinal axes were similar for the sternal posterior rotator, with subtle, significant differences along the dorsal-ventral axis ($p<10^{-7}$, Mann-Whitney U test). The synapse prediction algorithm found ~50% more synapses for that neuron, and this was accompanied by a small increase in the density of predicted synapses near the peak of the distribution. When we examined synapse predictions in that plane, we found small branches that were missed by manual tracking. These also appeared to be regions with noticeably higher concentrations of synapses compared to more ventral regions.

We conclude that comparing manual annotation with automatic synapse prediction in these synapse dense reconstructions includes four sources of noise (Schneider-Mizell et al. 2016): 1) completeness of manual tracing of fine branches, 2) completeness of manual synapse annotation, 3) completeness/accuracy of segmentation, and 4) accuracy of synapse prediction. For our ground-truth comparisons, we have ensured that all the manually traced dendrites are proofread and reattached to the segmented reconstructions, thus limiting the third source of noise. In Figure 2F, we show that for the four neurons we examine here, most of the synapses are from a small number of partners that make many synapses, and those partners are largely correctly identified from both the manual and predicted synapses. Due to the massive effort required to manually annotate synapses even to this degree, we recommend that users take the automatic segmentation at face value and visually inspect synapses when precise counts are necessary. We have added an Extended Data Figure 3 to present this analysis in more detail, and we identify these noise sources in the figure legend.

Fig 4 and 5 : link to the synapses or some EM images in the figure would illustrate the type of synapses identified.

We now include several example EM images of synapses in Extended Data Figure 3.

mat and methods : a description of the synapse annotation/proofreading is missing

We now include a more complete description of the basic proofreading procedure. We also include a brief description of the synapse prediction methods described in Buhmann et al 2021.

Fig S4 : legend for panel E missing

We now provide a legend for panel E in Extended Data Figure 5.

Referee #2 (Remarks to the Author):

The manuscript, “Tools for comprehensive reconstruction and analysis of Drosophila motor circuits” by Azevedo, Lesser, Mark, and Phelps et al., presents a draft connectome of the Drosophila ventral nerve cord (VNC). This is based on an automated segmentation by zetta.ai of the FANC dataset. Additionally, a proofreading platform based on FlyWire is released, and segmentation and platform will enable researchers to fully proofread the FANC dataset in the coming years. Azevedo et al. then utilize these data, along with datasets of different modalities, specifically sparse genetic driver lines and x-ray data, to study which VNC motoneurons innervate which leg and wing muscles.

Major

The manuscript's focus seems to oscillate between the release of a novel dataset resource (‘the yet-to-be-proofread FANC segmentation’, Fig1 and Fig2), and a study about motor neuron muscle innervation (Fig3, 4, 5, 6). The unveiling of a non-proofread segmentation/connectome of an EM dataset, which was published a while back, doesn't seem groundbreaking unless the connectome's usefulness without proofreading is clarified. This means the novelty and discovery of the studied biological questions bear significant weight.

We agree with the reviewer that our paper has two significant contributions, both of which build on a serial-section EM dataset that was previously published. The first advance we report in this study is the automated segmentation of the EM data from Phelps et al. 2021. We report the first application of these segmentation techniques to the VNC, which drastically accelerated the utility of the VNC connectome by the scientific community. To assess the quality of the segmentation and synapse prediction, we directly compared the synapse detection algorithm against ground truth for 4 large motor neurons. To our knowledge, this is the first comparison of this kind.

The second advance is to identify the muscle targets of each motor neuron controlling a leg and wing. While this information is now effectively a resource that accompanies our dataset, it was important to understand the logic of upstream wiring and required significant innovation and investigation. To make these distinct contributions more clear, we have changed the text where we transition from describing the segmentation and synapse detection to describing the motor neuron identification, in the intro (line 69) and in the results (line 131 and 136).

Over 6000 neurons have already been proofread and proofreading in FANC is an ongoing community effort that is making rapid and steady progress. We chose to submit this paper prior to having achieved comprehensive proofreading to facilitate the dissemination and use of the dataset and associated resources by the scientific community. We have revised the discussion section of the text to communicate the rationale behind the community-driven proofreading: “The community effort to comprehensively reconstruct neurons and measure synaptic connections within FANC is ongoing. At the time of publication, the community has proofread less than half of the total FANC dataset, yet the automated reconstruction, proofreading environment, and analysis tools already enable interactive inquiry into the neurobiology of the VNC. Our work parallels similar efforts in the adult fly brain, the male VNC and the larval nervous system. Further, the distributed nature of the community-driven FANC effort is a strength, as it allows the connectome to be continuously refined as each circuit is proofread by experts in that area. Even before the connectome is complete, researchers can analyze subcircuits as a means of generating and falsifying hypotheses.” (line 248).

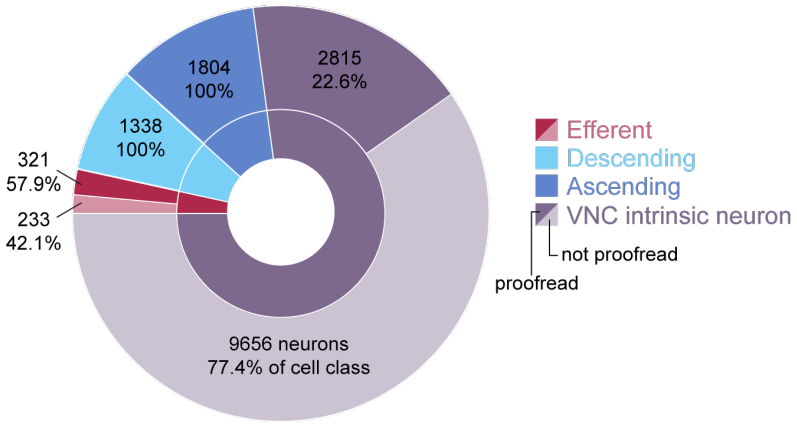
Our submission of this paper was intended to expedite the publication of other papers from the community, and to attract new users to FANC. Currently, five manuscripts are on the bioRxiv preprint server that include analyses of FANC, including our manuscript quantifying the organization of synaptic input to leg and wing motor neurons (Sapkal et al. 2023; Yang et al. 2023; Lesser et al. 2023; Dallmann et al. 2023; Cheong et al. 2023). These preprints already demonstrate the utility of this connectome for scientific discovery. (An up to date list of publications using the FANC connectome can be found here:

https://github.com/htem/FANC_auto_recon/wiki/#publications-and-public-data-releases)

The authors assembled an impressive consortium of 26 labs for VNC proofreading, underscoring the necessity of this process. However, there's no estimate of how long it will take to proofread the entire VNC, nor is there any statistical data on the consortium's monthly contributions or the time taken to correct major errors for neurons, beyond numbers for very few cells. More detail on this topic is required.

The reviewer makes an excellent suggestion. We have now quantified the collective proofreading effort to find that about 323 edits are made per day by 145 separate researchers from 30 difference labs. The 1,636 premotor and motor neurons that we have focused on required an average of ~20 edits per cell. At the current rate of 117,895 edits per year, with an average of 20 edits per neuron, for the 8,000 VNC intrinsic neurons left to be proofread, we’d predict ~1.4 years to complete proofreading. However, there are several factors make such estimates challenging. For example, the rate of overall proofreading is accelerating as more people continue to work in the dataset. There is also the possibility that “professional” proofreaders from the FlyWire consortium will contribute to the proofreading effort in the near future. These and other factors make it challenging to accurately estimate how long it will take to proofread all neurons in the FANC dataset.

Below is a summary of proofreading efforts so far, broken down by cell class. We removed sensory neurons because we do not yet have an accurate estimate of the total number; we currently have 3,700 sensory axons, but that is still growing. Sensory neurons are among the most difficult cells to segment and proofread, yet an intense focus for both our labs and others. Additionally, this quantification of proofreading progress is an underestimation, due to the distributed organization of the FANC community. We are in the process of building a robust central database to keep track of neurons that have been coarsely and finely proofread, but currently we are still reliant on individuals keeping track and uploading neurons they have proofread to a central database. We include a plot of the current status below, but we have chosen not to include it in the manuscript, as it is an underestimate and will quickly be outdated. We could add an updated plot to the revised manuscript if the Reviewer feels that would be useful.



Although the authors heavily emphasize the quality of the automated reconstructions, they conclude that only automated segmentation supplemented with appropriate proofreading, provides a satisfactory level of accuracy for circuit analysis. The level of proofreading required should be dictated by the question at hand. It might be more beneficial to apply "error bars" on the analyses in Fig3, 4, 5, and 6, instead of starting this topic broadly. Was proofreading essential to arrive at these conclusions later, at all?

We agree that the level of proofreading should be dictated by the scientific question. In some cases, cell type identity was possible without proofreading the segmentation. To answer the reviewer's question, yes, proofreading was absolutely essential to be able to identify the MNs and analyze their connectivity. In a new Extended Data Figure 6, we quantify the effect of proofreading by comparing the volume of motor neuron objects in the initial segmentation to the proofread MNs. Many MNs were initially missing identifying branches, and other objects were initially attached to glia.

To address the reviewer's comment more broadly, every neuron in FANC and other EM datasets requires manual inspection to assess its completeness prior to analyzing anatomy or connectivity. In some lucky cases, "appropriate proofreading" might be just a simple inspection. In intermediate cases, only a couple edits (merges or splits) would be needed to get a cell into decent shape. For example, some primary branches

of leg motor neurons crossed knife marks, so we were unable to match them to light-level morphology from the automatic segmentation alone. In more extreme cases, a user may need to spend hours doing fine proofreading of a neuron's dendrites, for example if they care about dendritic integration within that cell. The need to tailor proofreading granularity to the scientific question is a key reason why we feel that community-led reconstruction is superior to the alternative approach of proofreading up to some fixed threshold of moderate completeness, which some other groups have taken. In the FANC dataset, future users interested in building fine-scale compartmental models of leg motor neurons to study dendritic integration can further improve the proofreading we have already done. We have revised the text to explain our argument about community-led proofreading (line 248), and we clarify proofreading procedures by expanding on the "proofreading" section in the Methods (line 568).

"Manual" visual comparison appears to be key in matching MN identity across the imaging modalities, which could be a potential weakness. Is there a way to make this process more quantitative? What is the degree of similarity between the next best possible matches?

The reviewer is correct that visually comparing light microscopy (LM) images of MNs with EM reconstructions in FANC was key to assigning muscle targets. However, visual comparison was not the first or only criteria. To better explain the process, we have added text from the Supplementary Methods (previously the Appendix) to the Methods (line 609). Briefly, we first compared easy-to-characterize features, like whether the soma was anterior or posterior of the neuropil, and which of the four nerves the axon exits. Then, we used a computational analysis from Phelps et al. (2021), which relied on the NBLAST algorithm (Costa et al. 2016) to identify bundles, or groups of neurons with primary neurites that run together from the soma to the axon. Phelps et al. limited this analysis to the primary neurites of the manually-traced MNs. As a result, a large bundle in the leg nerve included neurons that could not be distinguished from one another on the basis of primary neurite tract alone, whereas the EM reconstructions revealed characteristic dendritic morphology. Next, we used these dendritic features to distinguish specific groups, such as the long medial dendrites of the long-tendon MNs or the tergotrochanter MNs. Careful visual inspection was the final step, at which point specific MN morphologies were readily identifiable. As we state in the Supplementary Methods, we did not attempt to trace LM images of MNs and then use a computational tool such as NBLAST.

To address the reviewer's concern, we developed an independent quantitative analysis that supports our claim that we can identify the muscle targets of each MN (Figure A18, also included below). The analysis assumes nothing about the MN morphology or any of the anatomical details we used for light/EM matching. Instead, we divided the neuropil into 8 um x 8 um x 8 um voxels. For each MN, we counted the number of synapses in each voxel (Panel A). We then normalized each vector and reduced the dimensionality of the vectors from 1891-D voxel space to 2D using the UMAP algorithm. This revealed clear clusters of MNs that aligned with our MN identification from comparing X-Ray, EM, and light microscopic data. Importantly, the number of MNs in each group matches the number of MN axons in the XNH data that

leave each nerve and innervate the matched muscles. The figure legend for Figure A18 provides additional details.

We also applied NBLAST to the reconstructed FANC MNs (Figure A18E). NBLAST values for neurons innervating the same muscle tend to be higher than across muscles, but hierarchical clustering based on NBLAST values did not give the same clusters (not shown). This is likely because most MNs share some morphological features, such as primary neurites that run in common tracts. This suggests that if we had traced LM images of neurons and relied exclusively on NBLAST to compare with FANC reconstructions, we would not have arrived at the same clusters or inferred the same muscle targets for MNs that belong to the same clusters. This provides an additional justification of our approach for MN identification.

We have added a paragraph to the Methods to point to this new analysis and to point the reader to more details in the Supplementary Methods (line 618).

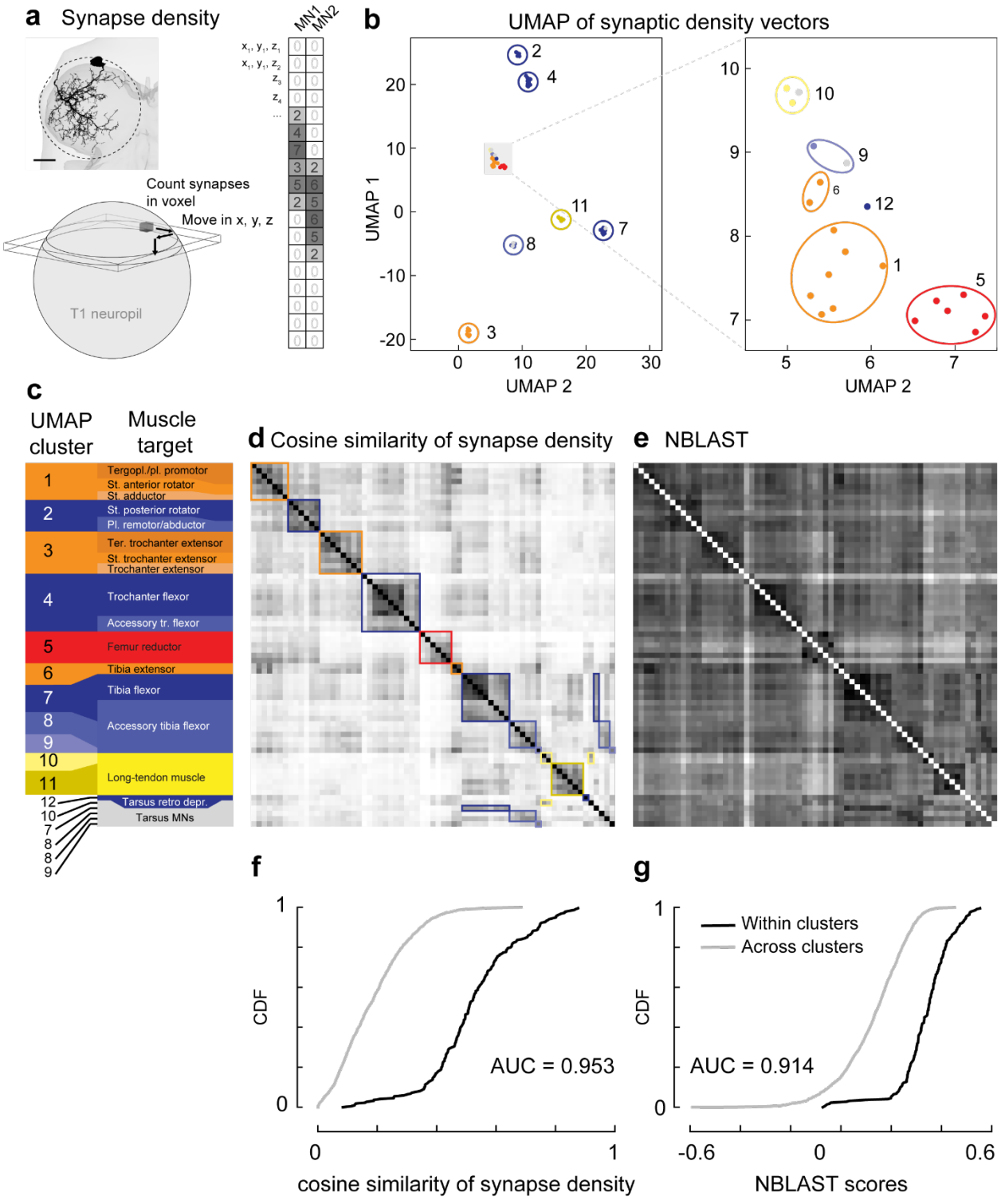


Figure A18. UMAP algorithm identifies clusters of MNs that align with our proposed MN identities. (**A**) Automatic prediction of synapse location allows us to compare the morphology of MNs by comparing the density of their input synapses in the T1 neuromere. We divided the neuropil into 8 µm x 8 µm x 8 µm

voxels and counted the number of input synapses in each voxel for each MN. (**B**) The 69 MN vectors of synapse counts were normalized and then embedded from 1891-D voxel-space into two dimensions using the UMAP algorithm (McInnes et al., 2020). L2 normalization spread the clusters out from each other compared to L1 normalization. Most MN synapse density vectors fell into distinct clusters. Several MNs embedded near one another (inset). (**C**) The clusters identified by UMAP are aligned with the MN muscle targets. (**D**) An alternative distance metric to the UMAP embedding is the dot product of the synapse density vectors, called the cosine similarity. We used scikit-learn to compute the cosine similarity (Pedregosa et al., 2011). The matrix of pair-wise comparisons shows higher pair-wise similarity within the clusters identified by UMAP. (**E**) NBLAST is a standard method for computationally comparing neuron morphologies (Costa et al., 2016). The pairwise NBLAST scores are also higher within the UMAP clusters. (**F**) The cumulative density functions (CDFs) of cosine similarity comparisons within (black line) vs. across (gray line) clusters. (**G**) CDFs of NBLAST comparisons within (black line) vs. across (gray line) clusters. The area under the curve (AUC) is the Mann-Whitney U statistic divided by the product of numbers of elements. This measures the discriminability of within vs. across comparisons. The AUC is higher for the cosine similarity of the synapse density vectors than for NBLAST comparisons.

<u>Minor</u>

The claim that "studying the compact motor circuits of Drosophila could identify circuit mechanisms conserved across limbed animals" seems overly optimistic. The insect VNC appears to be quite diverse and evolutionarily optimized for a particular species (Niven et al., 2008).

We have changed the text to read, "Thus, studying the compact motor circuits of *Drosophila* has the potential to identify general circuit mechanisms with relevance for understanding motor control in other limbed animals." (line 46) We agree that is best to avoid speculation about whether specific mechanisms are evolutionarily conserved vs convergent solutions.

The statement that "our approach of using multiple imaging modalities at different spatial scales can also be used to map the inputs and outputs of connectomes across species" seems overly broad. This might be true for insect species, but I am doubtful that this approach will work for mammals.

We have changed the final sentence of the introduction to, "Together, the reconstruction of a female VNC and the identification of the peripheral targets of MNs provide a foundation for investigating how sensorimotor circuits move the body and integrate descending signals from the brain to control behavior." (line 68). This phrase restates our two major contributions in this work, and while still broad, avoids the vague construction in the previous sentence.

The reference to the study on the adult fly brain (Galili et al., 2022) should be updated to include more recent publications, now that the connectome has been released.

Thank you. We now include references to the recent preprints accompanying the release of the Male Adult Nerve Cord and the annotated Full Adult Female Brain (Flywire) to the community (lines 59, 160, 201, 240, etc.).

I suggest a more detailed discussion on what other questions, beyond motor control, a proofread VNC dataset might be interesting for. This would underscore the dataset's usefulness beyond this study.

We agree with the reviewer that the VNC connectome will be useful for addressing questions beyond sensorimotor control. We have revised the discussion to state, “The existence of multiple VNC connectomes will enable comparison of neural wiring across individuals and the identification of sexually dimorphic neural circuits.” (line 240). In the final paragraph of the paper, we discuss how we envision FANC being used beyond understanding motor control, and outline the efforts required to make that possible (line 278). As we mentioned above, those will likely be contributions from other groups that have a stake in making sure that sensory neurons, ascending neurons, descending neurons and abdominal projections in FANC are accurately proofread and insightfully analyzed (line 240). Matching ascending and descending neurons up with their processes in the brain appears to be challenging enough that it may be simpler and faster to instead take the next technological step to section an entire CNS, i.e. a connected brain and VNC.

Referee #3 (Remarks to the Author):

The present manuscript describes both tools for circuit mapping and analysis and a specific analysis on the coordination of wing and leg motion when initiating flight. The methods include the Popvych et al. toolkit for alignment, Macrina et al. for segmentation, and Buhmann et al. for synapse detection, and Allen Brain Institute's tools for handling large volumes, proofreading with Neuroglancer. All of these tools are state of the art. A consortium of 26 labs is now proofreading neurons in the nerve cord of the fly, an indication of the impact that this work is already having, and which will eventually be harvested in the form of newly reported circuit maps describing the function of the nerve cord.

Outstanding that the team recovered between 60 and 85% of all cable for neurons of interest, compared to manual reconstruction, for the 4 example motorneurons. The number of synapses is within single-digit percent differences (Figure 2).

Figure 2 also states that "we used a synapse threshold of 5 to determine the total population of synaptic partners to compare." But this threshold, even if it reads sensible to me as an insider in the field, isn't justified in the text.

Thank you for pointing this out. Hulse et al. (2021) used a synapse threshold of at least 3, which we have now adopted for consistently. We feel this is a sensible choice from the literature to reject weak connections that may represent wiring noise. We have rerun the analysis with a threshold of three synapses. At this threshold, the left front accessory tibia flexor receives input from 65 partners according to the automatic segmentation and only 39 partners according to ground-truth. 40 and 28 of those partners make 4 or more connections. This implies that the fall off in the precision and recall plots is at the point when synapse partners with 3 connections are predicted that are not part of ground-truth (precision), or are ground-truth and are not predicted (recalled). We have changed Figure 2 accordingly. In another paper about premotor circuits in FANC (Lesser et al., 2023), we use a three synapse threshold to plot the matrix of synapse counts from presynaptic patterns onto MNs.

For quality assessment, 4 motorneurons were traced manually and then compared with the automatic segmentation. The differences and errors found are primarily of omission and not proportionally numerous, and the overall overlap is large. This sets a good basis for the work, assuming all other neurons are reconstructed to a similar level of accuracy where 60 to 85% of the cable, and most synapses, are recovered.

We appreciate the reviewer’s perspective. As the reviewer states, the errors that are visible in Figure 2 are primarily errors of omission, when the segmentation split branches that crossed image artifacts in the EM data. For our ground truth comparison, we selected neurons that were not affected by large merge errors. Motor neuron axons in the nerve were often falsely attached to glia that fill the interstitial spaces within the neuropil, creating complex merge errors. False glia merges were the hardest proofreading errors to fix, requiring the proofreader to carefully search for the one or more supervoxels within this snarl to split. These errors are evident when we compare the MN volumes before proofreading and after proofreading as a ratio greater than 1 (Extended Data Figure 6A). We plan to keep these examples in the figure to illustrate representative segmentation errors, but errors of omission like those we show were corrected, particularly for motor neurons that had been manually traced (Phelps et al., 2021). We have also added an additional figure illustrating how proofreading affected the synaptic input to motor neurons, as mentioned above in response to Reviewer 2’s comments (Extended Data Figure 6).

The comparison of 200 hours of manual reconstruction of a neuron versus 2 hours of manual proofreading of an automatically segmented neuron reveals just how far the field has gone. That these two neurons differ only in small ways that are irrelevant for the overall circuit architecture offers a reassuring view of the future of connectomics. Now if only these two hours of work could be automated away altogether; but that's beyond the scope of this paper.

We agree! We note though that many of the segmentation errors were due to large scale image artifacts or to tissue damage from the dissection. We are learning from that experience to improve tissue dissection and sectioning, which will also improve the segmentation. In addition, we have used the artifacts to train the next segmentation networks.

Figure 2F and 2G show a curious difference between the amount of precision and recall in the reconstruction of the 4 example motorneurons. For two of them, precision and recall are almost perfect; for the other two, it's fairly poor. Why is that? Surprisingly, the neurons with less synapses (as per Figure 2E) seems to have worse automated reconstruction. What explains this? Was the quality of the manual segmentation poor for these neurons and therefore the discrepancy is higher?

The reviewer makes an insightful point and we have reexamined the analysis in Figure 2 and updated it in the figure. The updated analysis shows that using either manual or automatic segmentation recovers nearly the identical partners for each MN, for partners making more than 3 synapses.

We realized that the prior analysis confounded several potential sources of error. We were attempting to test the quality of the synapse detection by comparing the presynaptic partners found with the manually annotated synapses vs. the automatically predicted synapses. However, it also assumed that presynaptic partners are all well proofread, while in fact the proofreading status of premotor neurons has improved significantly since these plots were initially created, though more for the left neuropil than the right. The left and right accessory tibia flexors that we show are small neurons that receive small numbers of synapses from only ~20 partners, so incomplete proofreading could affect this analysis.

We also created Extended Data Figure 3 to compare manual synapse annotation to automatic synapse prediction in more depth. There, we identify the sources of uncertainty in this comparison. One source that appears to affect this comparison is the manual tracing and synapse annotation itself, especially in synapse rich regions where small dendrites are sometimes hard to notice. Going forward, automatic segmentation will make it possible to examine an entire neuron much more quickly than manual tracing, thus making it more likely that fewer dendrites are missed due to proofreading fatigue or other human errors.

Interesting that the number of motorneurons controlling a leg are in the order of single-digit numbers. This both enables their identification and assignment to GAL4 lines, and renders the system numerically simple enough that the field has a fair chance to understand how it operates. This ought to be highlighted in the Discussion, generating ammunition for study sections to support this work, as this is evidence of experimental and theoretical tractability leading to high bang for the buck, with implications for engineering disciplines seeking to recreate robotic systems, beyond neuroscience.

We appreciate the perspective and we agree. We now state in the discussion (line 281), "The compact sensorimotor circuits that mediate robust control of the fly leg and wing may provide inspiration for engineering of micro-scale robotic systems (Dallmann et al. 2023)."

Polyinnervation – multiple motorneurons innervating the same muscle fiber – is known to occur in juvenile mammals but not adults; here, it is found in the fly leg, just like it was reported as well for the fly larval body wall muscles before (see Matthias Langraff lab work).

The comparison to the larva is a good one, and we have added the citation suggested (Landgraf et al. 1997) (line 173). Polyinnervation of adult insect muscles has been described (Hoyle 1983, cited), but to our knowledge, exactly which muscle fibers could be expected to receive polyneural input was not.

In which other arthropods (line 182) do motorneurons present central output synapses? A citation is missing.

Thanks for catching this. This specific line was in reference to gap junctions in the stomatogastric ganglion, and we have cited Marder, Guttierrez, and Nusbaum 2017 for a recent review (line 185).

In the same paragraph, we have also added a citation to work showing that the tibia extensor MNs of the hind leg of the locust make chemical synapses directly onto the flexor MNs (line 181). This connection was not observed in the pro- or mesothoracic MNs in the locust, and appears to be related to loading force in the enormous extensor jump muscles before takeoff. We examined whether tibia extensor MNs in T1 and T3 in FANC make chemical connections onto other neurons, but we observed only false positives or proofreading errors. Examining predicted output synapses was actually an effective way of proofreading MNs.

Drosophila use the tergotrochanter (TT) MNs of the middle leg to jump and fly when escaping predators and looming stimuli. The TT MNs are electrically coupled to the Giant Fiber neurons that drive escape. We

have not been able to resolve gap junctions at the current resolution, but gridTape-ssTEM offers the ability to identify regions where we expect gap junctions, and return to that precise slice and reimage at higher resolution.

Figure 4: in the legend, 'C' reads like for panel 'B'. and 'B' for panel 'C'.

We have corrected Figure 4. The panel letters and legend are now correct.

The description of the Drosophila wing muscle motorneurons alone would have made an amazing paper; that the paper goes well beyond that renders it foundational, an instant classic. The text is accessible and readable, and the explanations on reconstruction methods and accuracy read more than sufficient to ensure the validity of the data as reference for future studies.

Thank you.

How do the motorneurons relate to the thoracic hemilineage-wise thermogenetic assays in Robin Harris et al. 2015 eLife?

Excellent question. Motor neurons derive primarily from two stem cell lineages, or hemilineages, called 15B and 24B. Harris and colleagues knew this and did not attempt to activate these hemilineages with thermogenetics. Instead, they focused on the hemilineages that produce the majority of the VNC interneurons. In a companion paper (Lesser et al., 2023), we identify the neurons in FANC that provide synaptic input to the motor neurons and classify them according to their hemilineages. We found that some hemilineages produce neurons that predominantly contact wing MNs, while others produce neurons that only contact leg MNs. Other hemilineages produce neurons that can contact either wing or leg MNs, but not both (a negligible set). This aligns with the findings of Harris et al. where thermogenetically activating a wing specific hemilineages drove various wing behaviors and activating leg specific hemilineages drove broad leg behaviors.

In Figure 6C, the volume of motorneurons is plotted against the number of synapses that they receive. Why the volume? Given proximo-distal tapering, and the position of synapses primarily at the distal tips, likely the better correlation is with amount of cable, particularly with the amount of cable with a Strahler number lower than 3. What motivated plotting against volume?

Throughout the drafting of this manuscript, we have often used the size principle of motor control as a framing device, since we also find a similar set of relationships in fly MN properties and recruitment during movement (Azevedo et al. 2020). This framing has waxed and waned in emphasis, but it motivated using the simple metric of volume. We have changed the text to return to the size principle again, to both motivate this comparison and to better link this paper to our forthcoming manuscript (line 226). It is also not trivial to measure the volume of a neuron, or the surface area for that matter, so we thought that this could showcase the capabilities of segmentation and computed meshes. Surface area, volume, as well as the detailed branching patterns of an MN, will all affect the intrinsic excitability. The electrotonic properties can in principle be modeled, but we do not yet know enough about the type, amount or distribution of voltage-gated, ligand-gated, peptidergic or neuromodulatory channels or inputs to motor neurons to attempt a full-scale biophysical model of the MNs.

The first sentence of the discussion needs work: that's not the utility of the connectome but its very definition. The best description of the utility of the connectome I've ever read is in Denk et al. 2012 where the authors write the connectome is great to disambiguate between competing hypotheses of neural function.

The first two paragraphs of the discussion can be dispensed with, as they rehash results and not very well.

Thank you for the feedback. We have excised and rewritten parts of the Discussion for concision and emphasis.

If you are to discuss connectomes in the context of linking neural dynamics with motor output and posture, don't miss Heckscher et al. 2015 Neuron.

Excellent suggestion. We now state in the discussion (line 282), “Matching cell types and comparing connectivity motifs to the larval connectome may provide insight into the development and evolution of sensorimotor circuits (Heckscher et al. 2015).”

The larval connectome isn't complete yet: only for the whole brain and SEZ, and for select segments and subsystems of the nerve cord. If the last sentence of the discussion refers to comparisons of the motor systems of the larva and the adult, then indeed such comparisons are now possible.

Thank you, we have changed the sentence to read, “Matching cell types and comparing connectivity motifs to the larval connectome[62,81,102] may provide insight into the development and evolution of sensorimotor circuits[103,104]”

On Methods, line 314: the correct citation for CATMAID's tracing tools is Schneider-Mizell et al. 2016 eLife. The early paper by Saalfeld et al. 2009 introduced the image browsing plus text annotations only.

Thank you, we have modified the citation in both the Figure 2 legend and the text.

The Methods seem overly short, despite the appendix on identifying leg motor neurons. In particular, if another team of researchers wanted to reproduce the results, I am not confident they could merely from the description. The softwares mentioned are complex and take a lot of work to get up and running, and there are far more parameters than the subset mentioned here. Attaching configuration and setup files, together with links to software repositories and IDs of the specific versions used for the various softwares used could go a long way to address this concern.

We have significantly expanded the methods section and now provide a Github repository with information about the specific versions of software used for FANC segmentation and proofreading tools. https://github.com/EllenLesser/Azevedo_Lesser_Phelps_Mark_2023/

On the quality checks mentioned in the methods: if these were done using custom scripts and neurons encoded in CVS files or JSON files, why not attach them here. Would enable verification of the claims. The description provided is superficial and does not enable verification.

We now provide a Github repository containing scripts and files documenting the soma identification and quality checks. We also provide scripts executing the groundtruth vs. automatic segmentation comparisons. We have added details about the segmentation parameters and configuration to a separate public repository (https://github.com/htem/FANC_auto_recon).

Reviewer Reports on the First Revision:

Referees' comments:

Referee #2 (Remarks to the Author):

The authors have successfully addressed all my concerns. I suggest they publish a regularly updated version of the proofreading status pie chart on their website. Great work!

Referee #3 (Remarks to the Author):

Having read through the whole rebuttal letter, I am satisfied with the answers provided by the authors. This excellent work is now even better written, more accessible, more complete, and more reproducible.