nature portfolio

Peer Review File

Sub-cellular population imaging tools reveal stable apical dendrites in hippocampal area CA3

Corresponding Author: Dr Jayeeta Basu

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this manuscript, the authors present an automated detection algorithm capable of identifying dendritic and somatic ROIs in dense datasets. The algorithm is based on the existing constrained non-negative matrix factorization (CNMF) method, with the main novel aspect being its initialization stage, which uses minimal morphological assumptions and can therefore simultaneously detect both somatic and dendritic ROIs. In addition, the authors offer some new approaches for the refinement of ROI estimates and screening calcium transients, such as assessing the skewness of the distribution of fluorescence values and using a 'fitness curve' to detect significant transients. The combined methodologies provide a useful analysis tool for simultaneously detecting somatic and dendritic ROIs in dense field-of-view (FoV) data.
Next, the authors applied their algorithm to different density datasets recorded in area CA3 of the hippocampus during a spatial navigation task. Their analysis of the spatial tuning properties of somata, basal and apical dendrites showed that, in familiar environments, the different subcellular compartments exhibit similar spatial selectivity, revealing two main differences: 1) basal dendrites exhibit lower event rates compared to CA3 somata; and 2) apical dendrites show higher levels of between-days place field stability compared to basal dendrites and somata. These findings suggest a functional divergence between subcellular compartments of CA3 neurons for long-term spatial representation.
The investigation of dense imaging fields can help deepen our understanding of neuronal populations, and the method introduced in this study has the potential to be a valuable tool for effectively analyzing subcellular compartments within these dense datasets. However, the main novelty of this method lies in its initialization stage, while the majority of the methodology is based on the previously published CNMF algorithm. I am not sure that this contribution alone is substantial enough for publication in Nature Communications.
In addition, the method is not fully validated. While manually marking dendritic ROIs in a dense FoV is a heroic effort, it is not a reliable ground truth, as even experts are bound to miss ROIs, make mistakes, have inconsistent judgements or are unable to judge ambiguous cases (e.g. crossing dendrites). Moreover, while the authors present convincing evidence that their method performs better than existing algorithms on their data, they do not explore the boundaries of its performance - how dense can the field be before the method no longer performs well? A possible approach to address these concerns is to leverage the sparse datasets included in the manuscript. In sparse datasets, the detection of ROIs can be achieved more exhaustively and reliably. By overlaying multiple sparse FoV movies and summing up their activity frame-by-frame to generate an artificial dense FOV movie, the authors could create dense datasets with well-defined ground truths. Gradually overlaying more sparse fields would enable the generation of ground truth fields with progressively higher densities, facilitating the testing of the algorithm's performance limits in ROI density. Simulated activity traces would also facilitate testing the performance limits in activity sparseness. These measures would be critical for demonstrating the efficacy of the algorithm in settings beyond the rodent hippocampus.
Furthermore, I have concerns about certain aspects of the method itself. First, in the initial step of the analysis the raw movies are downsampled to a temporal resolution of approximately 1.5 Hz. This could potentially render it impossible to detect ROIs based on localized dendritic activity, which can be fast, lasting 120-200 ms. As a result, detected dendritic ROIs might be biased towards those with larger, long-lasting transients. The utilization of the 'fitness curve' to screen for significant transients also raises potential issues. This approach assumes that the entire dendritic ROI is activated with each calcium event, essentially disregarding the possibility of localized dendritic activity. Such events would manifest as small peaks in the fitness curve and categorized as not significant. Together, these issues might introduce a bias in the detection of dendritic ROIs and significant transients, potentially leading to the omission of calcium events. This is accentuated by how

much the results depend on the significant transient method: utilizing the fitness curve detected half the number of events as the Z-scored method, and the finding that there is a lower firing rate in the basal dendrites is therefore dependent on the fitness method. The ground truth for event detection here is not perfect as discussed above, and with no other techniques corroborating these findings they are raised further into question.
The findings regarding the properties and stability of spatial tuning in CA3 somata, basal and apical dendrites are very interesting: no other study has identified varying degrees of representational drift across subcellular compartments. However, the concerns above cast doubt on their reliability. Specifically, the observed lower event rate in basal dendrites could potentially stem from the algorithm's inability to detect certain events accurately. The observation of the higher stability of the spatial tuning across days in apical dendrites also relies on the accurate detection of calcium transients.
Finally, the manuscript contains minor errors that would benefit from some attention. Some figure panels are missing, and some figure legends do not accurately reflect the figures themselves. There are instances where terms and acronyms are used before being accurately defined. In addition, some sentences could be phrased more clearly, such as references to "the area under the curve" without specifying which curve, and statements about "high skewness values" without clarifying the corresponding distribution. Addressing these small issues would make the manuscript easier to read.

Reviewer #2

(Remarks to the Author)
In this paper, Moore et al describe a modification of the CNMF algorithm for finding more dendritic ROIs in population calcium imaging of neuronal compartments. This paper does not address the main problem of dendritic imaging, which is the backpropagating action potential, and thus I have difficulty understanding who the method is targeted towards. Studies of dendritic activity are always done with respect to the somatic activity, because the backpropagating action potential typically swamps out the smaller and more rare dendritic spikes. Identifying dendritic spikes that are not just backpropagating APs is thus the main problem of dendritic imaging, and this must be resolved if the present method is to be useful. In many cell types, dendritic spikes are not even present, or may be exceedingly rare. Are the authors convinced that dendritic spikes exist in these particular recordings?

The authors also do not address the oversplitting problem, where multiple dendrites from the same neuron are split into distinct ROIs. Combining these ROIs is necessary for follow-up analyses, since otherwise the neurons with many oversplit dendrites are counted more times than the neurons with undersplit dendrites. An automated method to do this would be welcome, and benchmarks should be done with respect to fully connected neurons, not individual dendrite pieces.

What is shown instead in this paper is an improvement over previous methods in terms of the raw quantity of ROIs that are detected, which is good but not particularly exciting. The failure to address the backpropagating AP and the oversplitting of the dendrites is a problem for all existing algorithms, and something they all fail at. However, I would expect these problems to be resolved by a method that claims to be specially developed for dendrites.

Reviewer #3

(Remarks to the Author)
In this manuscript, the authors present an algorithm, d-NMF, to automatically detect somatic and dendritic regions of interest (ROIs) in a field of view (FOV), in which neuronal processes are densely labeled with a genetically encoded Ca2+ indicator (either GCaMP6 or 7b). Two-photon Ca2+ imaging datasets at varying labeling densities, obtained from hippocampal area CA3, an area challenging to access by imaging, are used to highlight the algorithm's properties. The nature of this specific preparation (pyramidal neurons in area CA3 (probably subarea a) can be imaged sidewise) allows tracking of large populations of dendrites from different compartments. Notably, this would not be the case for other brain areas. The algorithm's results are validated by comparison with human expert manual labeling. However, it is doubtful, in my opinion, that dense FOVs such as those shown in Extended Data Fig. 1 (FOVs 11-14) can be segmented accurately.

Dendritic Ca2+ imaging is commonly used to understand how the dendritic processing of synaptic inputs shapes an individual neuron's action potential (AP) output. Usually, these experiments are performed in sparsely labeled FOVs because this allows 1) tracking of the dendritic ROIs over time and 2) assigning dendritic branches to a soma unambiguously. While the paper addresses the first point appropriately, the second, and probably more essential, the issue is not addressed by d-NMF (as briefly discussed in the 'discussion' section). I am generally not convinced that imaging FOVs with densely labeled dendrites can provide valuable insights into dendritic integration and its role in single-neuron function. However, I see the benefits of using d-NMF in automatically segmented sparsely labeled dendritic fields (often, algorithms focused on somas fall short here). I would suggest emphasizing this aspect much more in the paper. In addition, evidence and discussion should be added to address the following issues:

1) Throughout the paper, critical basic experimental descriptors need to be included. For example, what is the path distance from the soma to the given imaged dendritic ROI? Is there a systematic difference in path distance between basal and apical dendrites imaged? Which region of CA3 was imaged?
2) How are ROIs defined? As far as I can tell, ROI detected with d-NMF can contain multiple branches, single branches, and areas that contain only part of a branch. That means that the nature of these Ca2+ signals will be a complete mix: it could be significant single-spine signals, local dendritic-spike associated signals, or it could be backpropagating APs. Can d-NMF address this problem?
3) Using skewness as a parameter to classify ROIs introduces a clear bias in ROI detection. It will also limit the cell types for which this algorithm can be used.

4) Along the same lines, the findings related to event rates and spatial selectivity in the different dendritic compartments need to be validated via imaging in sparsely labeled Ca2+ imaging datasets. For example, how can event rates in basal dendrites (a dendritic region where dendritic Ca2+ signals are known to be dominated by backpropagating action potentials) differ from somatic event rates, which are known to be driven by action potentials? Could this be a result of the variable ROI sizes and patterns?
5) The fitness trace is an appropriate parameter to quantify if all Ca2+ events in a given ROI involved the entire ROI. But this is not true. What about, for example, single-branch signals or NMDA spike-mediated Ca2+ signals? This method would reject them because not the entire ROI is involved. This introduces bias, which should be at least discussed.

Minor points:
1) Line 101: figure 1b does not have a single neuron labeled
2) Line 144: figure 2c does not exist; maybe this should be 3c?

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
We appreciate the great efforts the authors have made to address our concerns: many of the concerns have been addressed. In particular, the use of synthetic data to demonstrate how density affects the sensitivity of d-NMF and the use of a sparse preparation to validate their experimental findings are valuable additions to this study.
However, I am still concerned about the ability of d-NMF to detect localized dendritic transients. Clarifying this point is critical for future users, as well as validating the experimental results presented in the manuscript. Low SNR and fast timescale localized dendritic events have previously been suggested to play a role in the hippocampus in memory formation. It seems entirely plausible that SNR and event duration may vary as a function of the dendritic compartment (basal vs apical dendrites) and dendritic distance from the soma and that these may explain some of the CA3 results presented here.

I still find it difficult to understand how downsampling to a rate of 1.5 Hz does not hinder the ability to detect fast (~200 ms long) localized dendritic spikes(as well as the performance of the algorithm when combined with other, faster indicators). The authors have demonstrated that there were similar ROIs detected in the full session (not-downsampled) and the downsampled session (Extended Data Fig. 2); however, they have not demonstrated the constraints of SNR, event frequency, and event duration on their ability to detect localized transients.

One approach to demonstrate this is to leverage their techniques for synthetic data (used in Figure 3): artificially insert events into localized dendritic regions with a range of spatial extents, signal amplitudes, frequencies and durations and demonstrate the bounds by which the algorithm performs.
Similarly, extended Data Figure 3 demonstrates the ability to separate spatially contiguous dendrites that have varying levels of independence, but the SNR for dendritic compartments appears to be unreasonably high, and the degrees of independence for which the algorithm can successfully identify independent ROIs are not shown. Consider the (very realistic) case where the full dendritic branch is activated synchronously throughout a recording session, except for one or two localized (~ 10 um) events of small amplitude (as has been described in hippocampal neuron dendrites in vivo). Will d-NMF be able to identify such events and divide this branch into smaller ROIs? With the synthetic data in extended Figure 3, exploring these bounds should be straightforward.

Reviewer #2

(Remarks to the Author)
The authors claim that my comments were "orthogonal" to the goals of the manuscript and thus did not address them. This manuscript presents a lot of work and that is impressive in itself. The second half of the manuscript could even be presented separately as a scientific study (and maybe it should). However, I don't think that diminishes the concerns I have specifically about the computational method, and the need to critically evaluate this part *as* a computational method.

To repeat, the two main problems I see with such dendrite segmentation methods are the backpropagating AP and the oversplitting of ROIs. The authors do not address the backpropagating AP, and claim that oversplitting does not change statistical tests. The choice not to address my comments has directed me to look more carefully at the design of the study and I have identified two new issues. Not only is d-CNMF oversplitting ROIs, but the human annotators making the ground truth have themselves been directly instructed to oversplit ROIs, which is described in the methods:

"Additionally, labelers were instructed to break up long dendritic branches into smaller ROIs about the length of the soma (approximately 30 pixels)."

I don't think this is ok to do. Breaking the dendrites into pieces of approximately this length will make the ground truth more like some algorithms (presumably more like d-CNMF) but this will not constitute a meaningful ground truth comparison. IF the ground truth is biased in this way, then my second new issue arises: older methods were not optimized to extract "soma-

length" segments of dendrites. In fact, I believe both CNMF and Suite2p are trying to find segments as long as possible. These older algorithms have also not been optimized by the authors for the number of ROIs returned, even though some of their main parameters control this (number of clusters in CNMF and threshold in Suite2p).

Even though the authors did not directly present how much oversplitting from d-CNMF there is in the real data, we can infer it from the new analyses using simulations: at the 1.5 quality threshold with 50-100 neurons in the field of view, you are oversplitting neurons by a factor of 20-40 (!!) while at the more conservative 3.0 quality threshold it is still a factor of 5-10.

It cannot be mathematically true that oversplitting does not affect statistiscal analyses. If an analysis has multiple copies of the same replicates (5-20 copies of each replicate in this case), then the statistiscal tests will by definition be incorrect, since the assumptions of independence are strongly invalidated. No merging analysis can change this basic statistical fact. Further, the amount of merging for this analysis seems minor (only 30% reduction in number of apical dendrites, whereas their simulations show that at least 80% reduction is necessary, and sometimes 97% reduction depending on data quality and SNR). The authors could claim that the simulations are designed to push the limits of the method, but 50-100 neurons in a field of view seems relatively normal and the oversplitting seems even worse at lower neuron numbers. Still, the amount of oversplitting in real datasets can be easily gaged by running more sophisticated clustering methods, and visualized with dimensionality reduction methods, like tSNE and UMAP.

For investigating the effect of the BAP, you can evaluate whether any activity exists in local dendritic compartments that is significantly different from the rest of the dendritic tree. Do any of your ROIs look like spines? If there isn't any localized dendritic activity, then we're looking at different versions of the BAP in different parts of the dendritic tree. Which is fine and potentially interesting, but it does not really address questions about dendritic integration and computation.

If my concerns are not seen as relevant here, then I am not sure what the method is innovating on. Is this method really an improvement over existing approaches like Suite2p and CNMF, which also give you a lot of oversplit dendritic ROIs without considering the effect of the BAP, and which allow you to control how many ROIs you want to get? The new method is claimed to recover a higher fraction of ROIs, but the older methods were not using optimized parameters for these particular datasets. Simple changes such as increasing the number of components (CNMF) and reducing the threshold (Suite2p) would have increased the number of returned ROIs and thus the matching to the "ground-truth" ROIs, but these analyses were not done by the authors.

In summary, the authors have not addressed my original two concerns (oversplitting and BAP) and their comments led me to identify two new concerns (artificially over-segmented ground truth and failure to optimize previous algorithms to this artificial scenario).

Reviewer #3

(Remarks to the Author)
the authors have addressed all of my concerns.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have nicely addressed our previous comments. The manuscript presents an algorithm that has been optimized for defining ROIs over dendrites; there are novel initialization and refinement steps and I feel that many researchers in the community will be users. The authors also present new findings related to the functional properties of CA3 dendrites that I find interesting and compelling.

I have also been asked to comment on the most recent review from another reviewer:

Related to bAPs, oversplitting and how to merge dendritic segments: The authors argue that merging of ROIs is up to the user, which I largely agree with. It is not clear what the best method is to merge ROIs, and this process could be quite different for users with different types and quality of data. In fact, the underlying biology related to in vivo dendritic function is largely unknown, which further implies that it is premature to develop an algorithm that claims to have the correct merging procedure. This does not mean that an algorithm without such a capability is not useful. For example, with a particular merging algorithm implemented, it would be possible to look for differences in dendrites between conditions. Whether the algorithm is the “correct” one or not would, in such a case, not prevent authors from detecting interesting dendritic functional changes across conditions.

However, the authors do some merging analysis in their rebuttal using an off-the-shelf hierarchical algorithm, and I agree with the reviewer that this should be included in the manuscript to give readers an idea of how such a step could be accomplished. Such a section in the manuscript could be expanded and explored more with the synthetic data analysis that that authors have used so nicely up to this point. As presented, the authors use a merging threshold and demonstrate that their scientific results hold for a range of thresholding values. Such robustness analysis is valuable to the techniques

presented, but I believe that the reviewer's concern could be largely addressed by emphasizing in the text that such robustness analysis is necessary for users to perform on their own.

I also agree with the reviewer that newer comparisons to Suite2p should be included in the manuscript; and that the the benchmark (F1) is not necessarily a good one. Comparing to human annotators is also not necessarily good since there is arbitrariness here too. Here again, the authors could address these issues relatively easily with synthetic data.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
In this rebuttal, Moore et al clarified a few things for me. While some things got better, others did not change and others got worse. For my main criticisms in the previous rounds of review, the authors essentially answered by saying that dealing with the bAP and with the oversplitting of the dendrites is the user's problem. They illustrate how one might do that, which I appreciate. For example, they show individual examples of local and global activity and they show me "privately" a coarse-to-fine segmentation with an off-the-shelf hierarchical merging algorithm (but they leave this figure out of the manuscript). I am not sure what to do with this response. My comment was simply that I consider these two steps as *the* problem that needs to be solved for dendritic segmentation. Whether one gets 60% or 70% pixels covered by the dendrite segments is of little concern to the user. This paper is essentially trying to argue that the segments found by their algorithm are of a better quality in some sense than the ones that could have been found with non-dedicated algorithms. But for a full analysis, the users still need to merge the segments themselves, which can then be used to estimate global vs local activity. This merging step is crucial, not because we only want a single activity number for each neuron as the authors imply that I meant, but because we need to know which segments and which branches go together, so we can do the kinds of studies the authors enumerated in their reviewer table 1 which they made just for me. While the hierarchical merging algorithm is a start, it's a very basic approach, and there is no quantification of how well it works. I thus consider that the status of my two main previous concerns is unchanged.

Now for the things that got better: the authors showed that their scientific analyses do not depend on the level of merging, mainly because the analyses are done on an FOV basis. Further, they tell me (but not the reader), that this level of control analyses should be considered standard going further. Are you sure the users will never do statistics at the level of ROIs? Sometimes this is unavoidable scientifically, and also often I suspect the users will not know that this is not an OK thing to do.

Now for the things that got worse, and only got worse because I realized the true nature of the benchmarks performed here. The authors evaluate this ROI segmentation problem as if it is a semantic segmentation task. This is inappropriate, and tells us very little about how well any of the algorithms work. An algorithm would get the same score if all the ROIs are split into literally individual pixels, or if they are all combined into a single ROI for the entire field of view. I think the reason I missed this originally is that I could not conceive of making this choice for evaluating a clear instance segmentation problem. There are so many ways to do well on a semantic segmentation task, none of which are shown (like thresholding a locally-normalized intensity map or maximum projection, or a correlation map, or a skewness map, or a variance map etc). But that wouldn't really capture at all the thing we care about, which is that the segments should correspond to meaningful, individually separated parts of neurons. In light of the previous rounds of reviews, no benchmark would really capture this, since the human annotators essentially split these dendrites into arbitrary segments, just making sure they are not too long.

I also noticed that Suite2p did in fact "improve" to an F1 score of 0.68 with the threshold of 0.75, which was in fact the same as d-NMF. However, this figure was only shown to me, and not included in the paper. Suite2p-0.75 should become the default comparison to use in the main text. Interestingly, a difference between 0.59 and 0.68 was reported as significant between d-NMF and Suite2p in the main figure, but the same mean values came out insignificant when varying the threshold of Suite2p for my "private" figure. This is probably not a meaningful distinction. I don't think these scores are very meaningful in general, but I also don't understand why the authors would hide the fact that simply decreasing a threshold results in more pixels in the output, and that's the main determinant of performance by this metric.

It is at the discretion of the editor what to do with these comments, but I am afraid I cannot give a more positive review. I am sorry I missed some of these crucial details on the initial review, and I thank the authors for pointing them out in the rebuttal.

(Remarks on code availability)

We have formatted this response as follows:

Original remarks from the reviewers in black italics

Our responses in black

Quotations from the updated manuscript in blue

Remarks and responses are numbered by Query (Q1, Q2, Q3…) and Answer (A1, A2, A3…)

For points common to more than one reviewer, we have addressed them with similar text in each individual subsection. When a relevant figure is included it is only included once.

Reviewer #1

Q1. *In this manuscript, the authors present an automated detection algorithm capable of identifying dendritic and somatic ROIs in dense datasets. The algorithm is based on the existing constrained non-negative matrix factorization (CNMF) method, with the main novel aspect being its initialization stage, which uses minimal morphological assumptions and can, therefore, simultaneously detect both somatic and dendritic ROIs.*

In addition, the authors offer some new approaches for the refinement of ROI estimates and screening calcium transients, such as assessing the skewness of the distribution of fluorescence values and using a 'fitness curve' to detect significant transients.

The combined methodologies provide a useful analysis tool for simultaneously detecting somatic and dendritic ROIs in dense field-of-view (FoV) data.

A1. We thank the reviewer for recognizing the usefulness of these tools.

Q2. *Next, the authors applied their algorithm to different density datasets recorded in area CA3 of the hippocampus during a spatial navigation task. Their analysis of the spatial tuning properties of somata, basal and apical dendrites showed that, in familiar environments, the different subcellular compartments exhibit similar spatial selectivity, revealing two main differences: 1) basal dendrites exhibit lower event rates compared to CA3 somata; and 2) apical dendrites show higher levels of between-days place field stability compared to basal dendrites and somata. These findings suggest a functional divergence between subcellular compartments of CA3 neurons for long-term spatial representation.*

The investigation of dense imaging fields can help deepen our understanding of neuronal populations, and the method introduced in this study has the potential to be a valuable tool for effectively analyzing subcellular compartments within these dense datasets.

A2. We thank the reviewer for appreciating the importance of investigating dense imaging fields and the utility of our approaches.

Q3. *However, the main novelty of this method lies in its initialization stage, while the majority of the methodology is based on the previously published CNMF algorithm. I am not sure that this contribution alone is substantial enough for publication in Nature Communications.*

A3. Our algorithm and tool kit have made several advances over previously existing methods. We are sorry we did not explicitly draw attention to these in the previous version of the manuscript but now elaborate on these to emphasize the differences.

While our approach does utilize some of the core approaches of CNMF, there are many other differences, both during computation and in the "clean-up" phase of keeping valid ROIs and identifying

true calcium transients. A broad comparison of the algorithmic design choices between our method and other methods is summarized in **Table 1** and bullet-points below.

Aside from these differences, the initialization stage is indeed an important step for these types of matrix factorization problems, as NMF is a non-convex problem with many local minima (Lee and Seung, Nature 1999). So the advancement of finding suitable initialization for dendritic ROIs is non-trivial and of interest to a wide range of experimenters.

From the revised manuscript:

"The initialization stage is an important step for these types of matrix factorization problems, as NMF is a non-convex problem with many local minima[36]."

"Though our approach utilizes NMF as the core extraction algorithm, there are a few notable differences between our approach and other published methods. Our initialization method is based on localized coactivity, and the number of components within a patch is discovered in a data-driven manner rather than specified beforehand. CNMF imposes no spatial constraints on the ROIs, while we enforce ROIs to consist of contiguous pixels. CNMF constrains the temporal component by an autoregressive fit, while we have no constraint on the temporal components. This allows for the possibility of calcium transients with different dynamics to exist within a single ROI. Our advancement of finding a suitable initialization for dendritic ROIs is an important advance of interest to a wide-range of imaging modalities, including axons[18,44,45] and a variety of cell types such as astrocytes[46,47]."

	CaImAn (CNMF with Sparse NMF Initialization)	**suite2p**	**d-NMF**
Initialization	• Nonnegative double singular value decomposition initialization • # of components must be specified	• Local correlations • # of components is discovered	• Local co-activity • # of components is discovered
Spatial Constraints	• None	• None	• Contiguous pixels
Temporal Constraints	• Autoregressive fit	• None	• None
ROI Selection	• Space correlation[a] • SNR • CNN classifier[b]	• Skewness • Variance • Anatomical measures[b]	• Skewness • Size • Spatial fitness[a]
Overlap	• De-mixed according to NMF	• Discard pixels belonging to multiple ROIs	• De-mixed according to NMF

Table 1 | Comparison of method details between d-NMF, CaImAn, and suite2p

[a]The “space correlation” metric described in the CaImAn paper is similar to the Fitness trace measure used in this manuscript

[b]The CNN classifier described in the CaImAn paper classified ROIs as having a soma-like shape, with a similar goal as the anatomical measures described in the suite2p pre-print.

- During computation, we do not make assumptions about the dynamics of the calcium indicator. We do not constrain the temporal traces to match template calcium transients modeled by autoregressive dynamics. This allows for the possibility of calcium transients with different dynamics to exist within a single ROI.

- Post-processing, CaImAn utilizes 3 measures to evaluate and discard components: Spatial footprint consistency, Trace Signal-to-Noise Ratio (SNR), and a CNN-based classifier. The spatial footprint consistency measure is related to our fitness trace, and the SNR measure is similar to our use of the skewness. The classifier is trained to detect somatic ROIs and is not applicable to dendritic ROIs.

Q4. *In addition, the method is not fully validated. While manually marking dendritic ROIs in a dense FoV is a heroic effort, it is not a reliable ground truth, as even experts are bound to miss ROIs, make mistakes, have inconsistent judgements or are unable to judge ambiguous cases (e.g. crossing dendrites).*

Moreover, while the authors present convincing evidence that their method performs better than existing algorithms on their data, they do not explore the boundaries of its performance - how dense can the field be before the method no longer performs well? A possible approach to address these concerns is to leverage the sparse datasets included in the manuscript. In sparse datasets, the detection of ROIs can be achieved more exhaustively and reliably. By overlaying multiple sparse FoV movies and summing up their activity frame-by-frame to generate an artificial dense FOV movie, the authors could create dense datasets with well-defined ground truths. Gradually overlaying more sparse fields would enable the generation of ground truth fields with progressively higher densities, facilitating the testing of the algorithm’s performance limits in ROI density. Simulated activity traces would also facilitate testing the performance limits in activity sparseness. These measures would be critical for demonstrating the efficacy of the algorithm in settings beyond the rodent hippocampus.

A4. This is an incredibly insightful and useful suggestion. Using one of our sparsely labeled datasets, we constructed synthetic dense datasets with known ROI positions and temporal traces (**Figure 3**). We demonstrate that our approach maintains an F1 score of 0.8 (out of 1) for detecting ROIs in fields of view containing up to 200 neurons. This analysis bolsters our claim that we can accurately detect ROIs in densely labeled fields of view, and is a valuable addition to our study.

In addition, the synthetic data allows us to evaluate the number of usable ROIs in a wide range of labeling densities (**Figure 3e**). This will be an important tool for experimenters to find a “sweet spot” of labeling density suitable for specific research questions while maximizing experimental throughput.

From the revised manuscript:

“To further validate the accuracy of d-NMF, we constructed synthetic dense datasets using one of the sparsely labeled datasets (Fig. 3). This sparse dataset had known ROI boundaries and temporal traces, serving as a ground truth. By copying, rotating, and shifting the original movie, we constructed synthetic dense datasets with up to 200 neurons with a high degree of overlap between dendrites (see Methods). We found that as we added more duplicates, the quality of signals decreased, due to the addition of more background from other ROIs and additional neuropil (Fig. 3c). This had the result that adding additional neurons beyond a certain point no longer yielded detectable ROIs or even decreased the amount of ROIs with a high quality signal (Fig. 3d). Nevertheless, for ROIs with a signal quality

above 2 standard deviations, d-NMF maintained an F1 score of 0.8 out of 1, outperforming the output of suite2p, another open-source analysis package[29] (Fig. 3e). This serves as evidence that our method is sound and accurate at detecting ROIs from dense fields of view."

"It is important to note, however, that increased density does not come for free. When we constructed our synthetic datasets by duplicating and shifting a field of view with a single neuron, the overall signal quality decreased as more neurons were added. This indicates that, for a given expected morphology and distribution of cells, a "sweet spot" of labeling density may exist which maximizes the number of usable neurons while minimizing the number of animals needed."

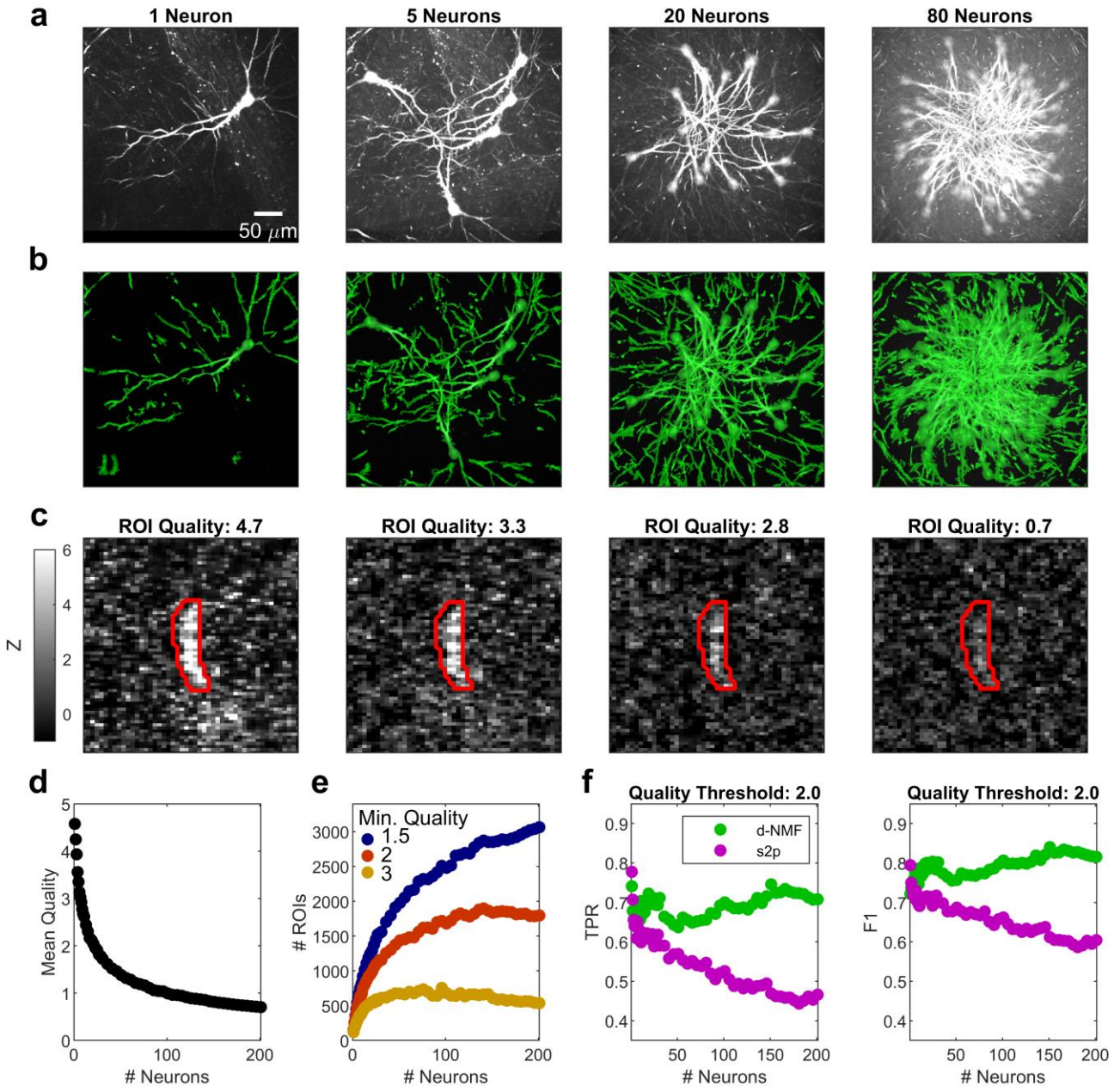


Figure 3 | Synthetic data demonstrates that d-NMF remains accurate at high density. a, Maximum projection images of movies generated by copying, rotating, and shifting the movie in the first column. **b**, ROIs derived from d-NMF for the corresponding movie in **a**. **c,** Frame of maximum activation for a sample ROI in the corresponding movies in **a**. The signal quality of this ROI is noted in the titles, and decreases as more neurons are added to the synthetic data. **d**, The mean signal quality of ROIs decreased as more neuron replicates were added to the synthetic data. Quality is defined as the mean z-scored value of all

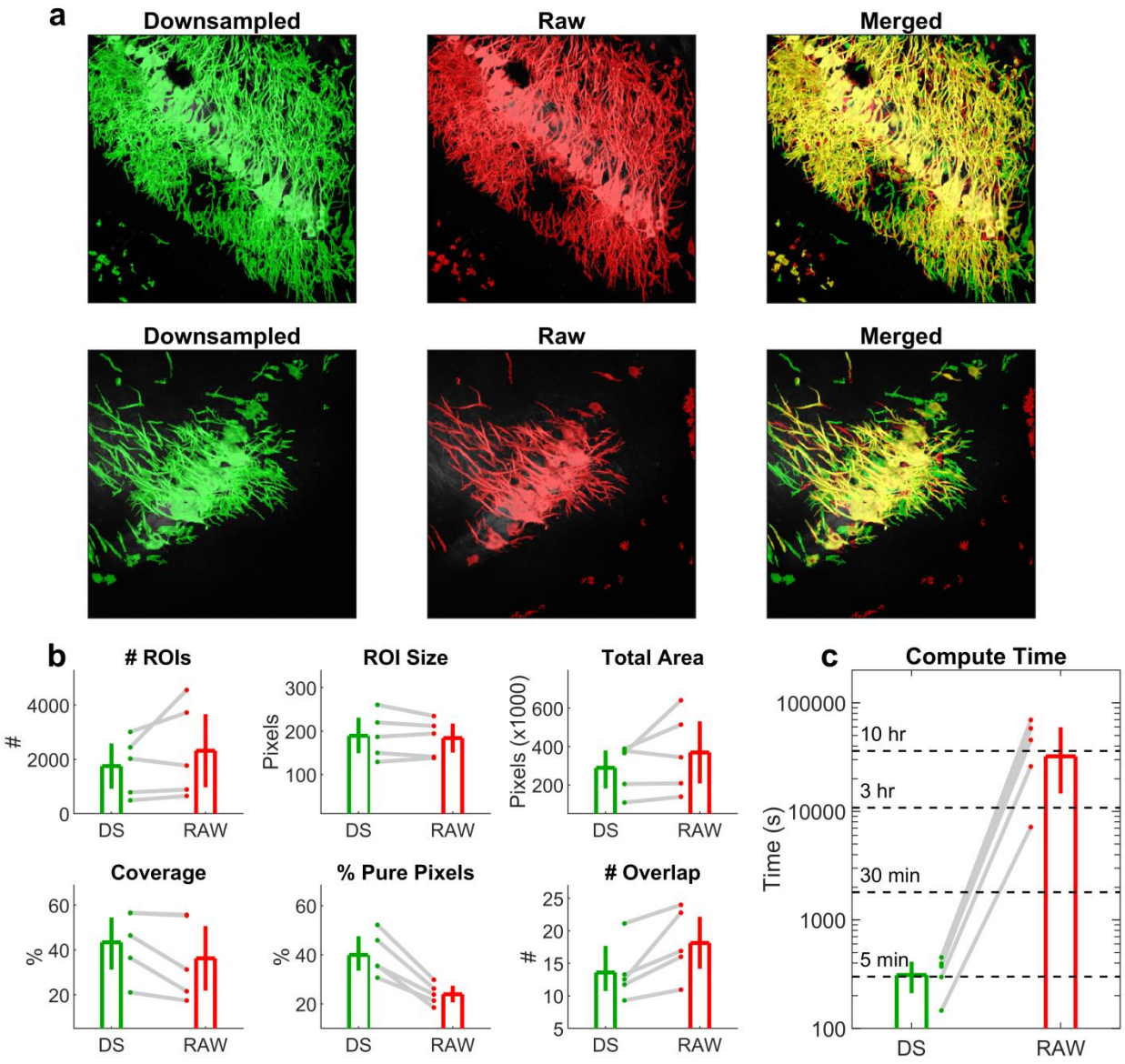
pixels at the ROI's most active frame. **e**, The number of ROIs with signal quality above a certain threshold (here 1.5z, 2z, and 3z) initially increased linearly but slowed down or decreased at higher densities. **f**, Left, restricted to ROIs with a signal quality >2z, true positive rate for d-NMF (green) and suite2p (s2p, purple) as a function of the number of neurons in the synthetic dataset. Right, F1 score for the same data. The d-NMF method maintained high performance even for very dense datasets.

Q5. *Furthermore, I have concerns about certain aspects of the method itself.*
First, in the initial step of the analysis the raw movies are downsampled to a temporal resolution of approximately 1.5 Hz. This could potentially render it impossible to detect ROIs based on localized dendritic activity, which can be fast, lasting 120-200 ms. As a result, detected dendritic ROIs might be biased towards those with larger, long-lasting transients.

A5. We re-detected ROIs in non-downsampled videos and found only minor differences (**Extended Data Figure 2**). Working with downsampled data reduced execution time by approximately 2 orders of magnitude, a dramatic savings in compute power and analysis time. Additionally, we were able to detect ROIs in the downsampled data that we could not in the non-downsampled data, likely due to a de-noising effect of averaging and downsampling.

From the revised manuscript:

"Temporal downsampling had little systematic effect on the ROIs detected, and actually detected ROIs in a larger proportion of the FOV, while saving two orders of magnitude of processing time (Extended Data Fig. 2)."



Extended Data Fig. 2 | Similar segmentation of down sampled and non-down sampled data. a, Two example fields of view (top and bottom rows) with ROIs detected from data downsampled to 1.5 Hz (Left, green), original data sampled at 30 Hz (Middle, red), and overlay (right). **b**, Many ROI properties were similar between downsampled (DS) and non-downsampled (RAW) datasets, including the number of ROIs detected (DS: 1700, [800, 2400]; RAW: 800, [500, 3200]; p= 0.84, Wilcoxon sign-rank test throughout), mean ROI size (DS: 160, [150, 220] pixels; RAW: 190, [140, 210] pixels; p=0.84), and total area of ROIs (DS: 260, [170, 380] pixels (x1000); RAW: 170, [90, 510] pixels (x1000); p=0.84). Total FOV coverage was significantly higher for the downsampled videos (DS: 38, [31, 56] %; RAW: 20, [15, 55] %; $p=7.8x10^{-3}$). The percentage of pixels belonging to a single ROI (DS: 38, [32, 52] %; RAW: 28, [21, 49] %; p=0.11) was also higher for the downsampled videos, but the difference was not statistically significant. The degree of overlap between ROIs was similar between downsampled and raw videos (DS: 12.6, [9.3, 20]; RAW: 13.5, [7.5, 22.8]; p=0.74). **c**, Execution time was 2 orders of magnitude longer when running d-NMF on raw 30 Hz data (51, [7.2, 89] seconds (x1000)) compared to downsampled, 1.5 Hz data (370, [200, 450] seconds). Each datapoint is a single FOV.

Q6. *The utilization of the 'fitness curve' to screen for significant transients also raises potential issues. This approach assumes that the entire dendritic ROI is activated with each calcium event, essentially disregarding the possibility of localized dendritic activity. Such events would manifest as small peaks in the fitness curve and categorized as not significant. Together, these issues might introduce a bias in the detection of dendritic ROIs and significant transients, potentially leading to the omission of calcium events.*

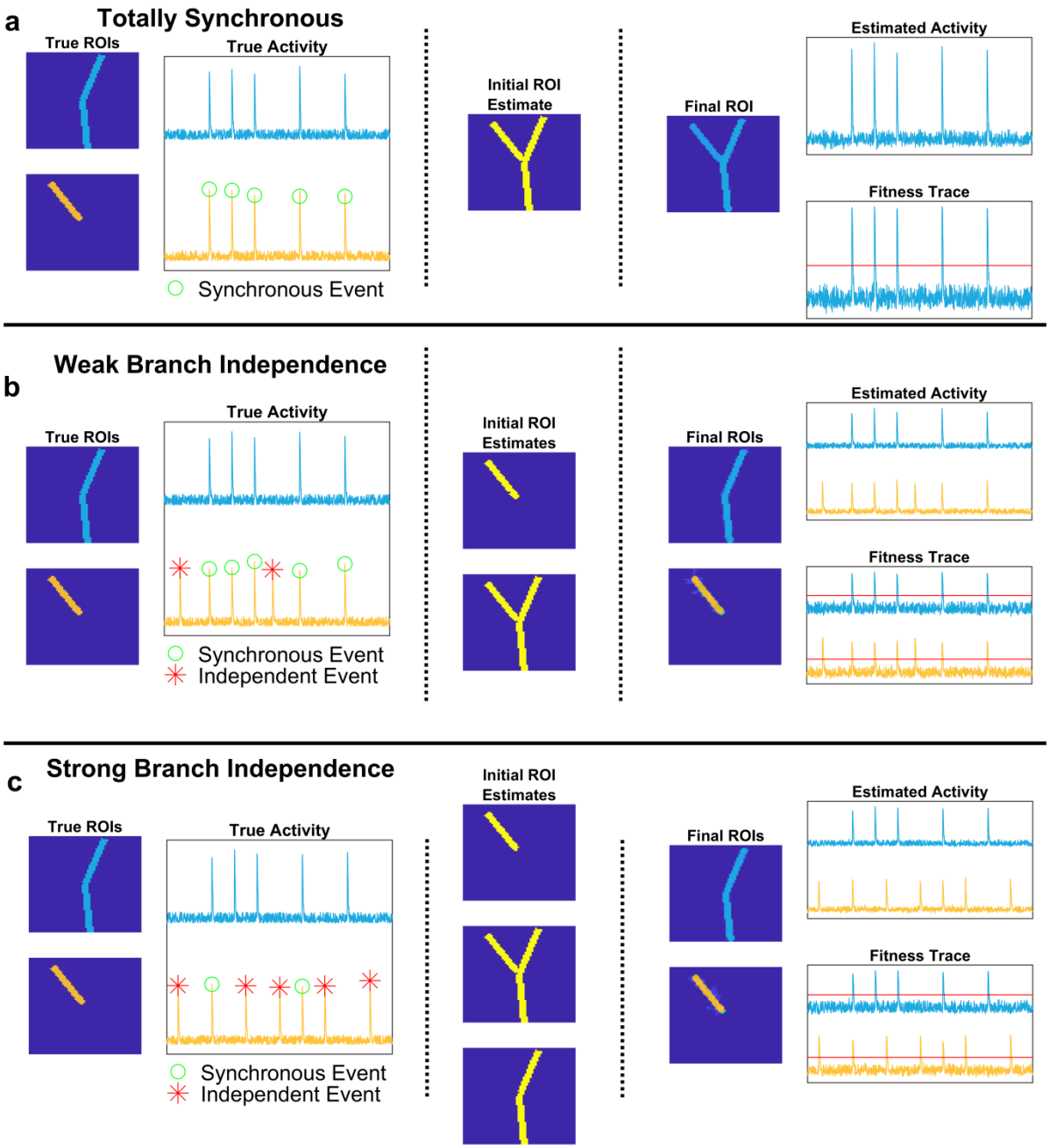
This is accentuated by how much the results depend on the significant transient method: utilizing the fitness curve detected half the number of events as the Z-scored method, and the finding that there is a lower firing rate in the basal dendrites is therefore dependent on the fitness method. The ground truth for event detection here is not perfect as discussed above, and with no other techniques corroborating these findings they are raised further into question.

A6a. We believe this is a result of unclear communication on our part. The fitness curve approach indeed requires the entire dendritic ROI to be activated for each event. However, this does not lead to a biasing of detection which would exclude fast localized events. This is because our algorithm defines ROIs as contiguous pixels that have synchronous activity. If one branch of a dendrite exhibits independent, localized activity, it will be treated as a separate ROI from its parent branch, and there will be no conflict. We have discussed and clarified this point using a new supplemental figure (**Extended Data Fig. 3**), which explores how the algorithm approaches 3 varying degrees of synchrony in a dendritic branch, from totally synchronous to heavily asynchronous. In all 3 cases the true activity of the branches was recovered.

From the revised manuscript:

" The resulting ROIs also have the property that their activity is homogenous within an ROI. Any dendritic branch or portion of a large ROI that has independent activity is split off into its own ROI (Extended Data Fig. 3)."

"An implicit requirement of the Fitness Trace is that the entire ROI must be active for a calcium transient to be detected, which could potentially discard localized dendritic activity and bias estimates. If ROIs are too broadly defined, such that they span multiple branches containing independent activity, this would lead to a bias to exclude branch-independent transients. However, the d-NMF algorithm identifies ROIs precisely as groups of pixels with synchronous activity. Dendritic branches that show independent activity are automatically split off from other branches and analyzed as a separate ROI (Extended Data Figure 3). In this way accurate ROI identification assists in accurate transient detection."



Extended Data Fig. 3 | Handling of branch-independent activity. a, Example synthetic dataset demonstrating synchronous activity within branches. Left, the true ROIs and activity are plotted. In this example, all events are synchronous between the two branches. Middle, the initial detected ROI cores estimate encompasses the entire tree. Right, after processing the estimated ROI still encompasses the entire tree, and the estimated activity is nearly identical to the true activity. All detected calcium transients cross the threshold for the fitness trace. **b**, Example situation where one branch is weakly independent from the other. Branch independent events are marked with red asterisks. The initial ROI estimates include the single branch as well as the entire tree. After the algorithm has converged, the remaining ROIs have separated the two branches and accurately estimated their activity. For both ROIs, no calcium transients are screened out by the fitness trace. **c**, Example situation where one branch is highly independent from the other two. The initial ROI estimates include the left branch, the right branch, and their union, but after processing the remaining ROIs represent the two branches separately.

A6b. Regarding the rate of detected events, it is absolutely true that the results can depend on the event detection criteria. **Main Figure 5**, which introduces the concept of the fitness trace, shows that using the fitness trace results in more accurate event identification than the 2z method (**Figure 5c**). We have expanded **Figure 5c** to include the actual event rate of manually-labeled events, and show that using the fitness trace results in an accurate estimate of the rate, and it is the 2Z method that (falsely) detects twice the number of events.

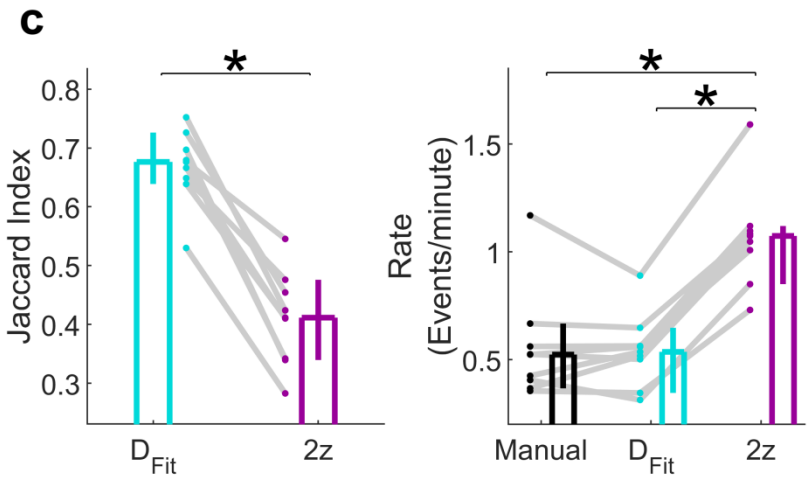


Fig. 5c. Left, classifier performance for DNMF with Fitness Trace (D_{Fit}) compared to a simpler detection method (2z, see Methods). D_{Fit}: 0.68, [0.64, 0.73]; 2z: 0.41, [0.34, 0.48]; D_{Fit} vs 2z: $p=3.9x10^{-3}$. Right, the detected event rate for the same 2 methods compared to manually classified events, illustrating the importance of correct classification. Manual: 0.52, [0.37, 0.67] Events/minute, D_{Fit}: 0.54, [0.35, 0.65] Events/minute; 2z: 1.07, [0.85, 1.12] Events/minute; n=9 sessions for all. Manual vs D_{Fit}: p=0.91; Manual vs 2z: $p=3.9x10^{-3}$; D_{Fit} vs 2z: $p=3.9x10^{-3}$. N=9 FOVs, Wilcoxon sign-rank test for all. Values reported as median and 95% confidence interval of the median.

Q7. *The findings regarding the properties and stability of spatial tuning in CA3 somata, basal and apical dendrites are very interesting: no other study has identified varying degrees of representational drift across subcellular compartments.*

A7. We thank the reviewer for recognizing the novelty and relevance of our findings pertaining to sub-compartment specific tuning features and differences in their stability.

Q8. *However, the concerns above cast doubt on their reliability. Specifically, the observed lower event rate in basal dendrites could potentially stem from the algorithm's inability to detect certain events accurately. The observation of the higher stability of the spatial tuning across days in apical dendrites also relies on the accurate detection of calcium transients.*

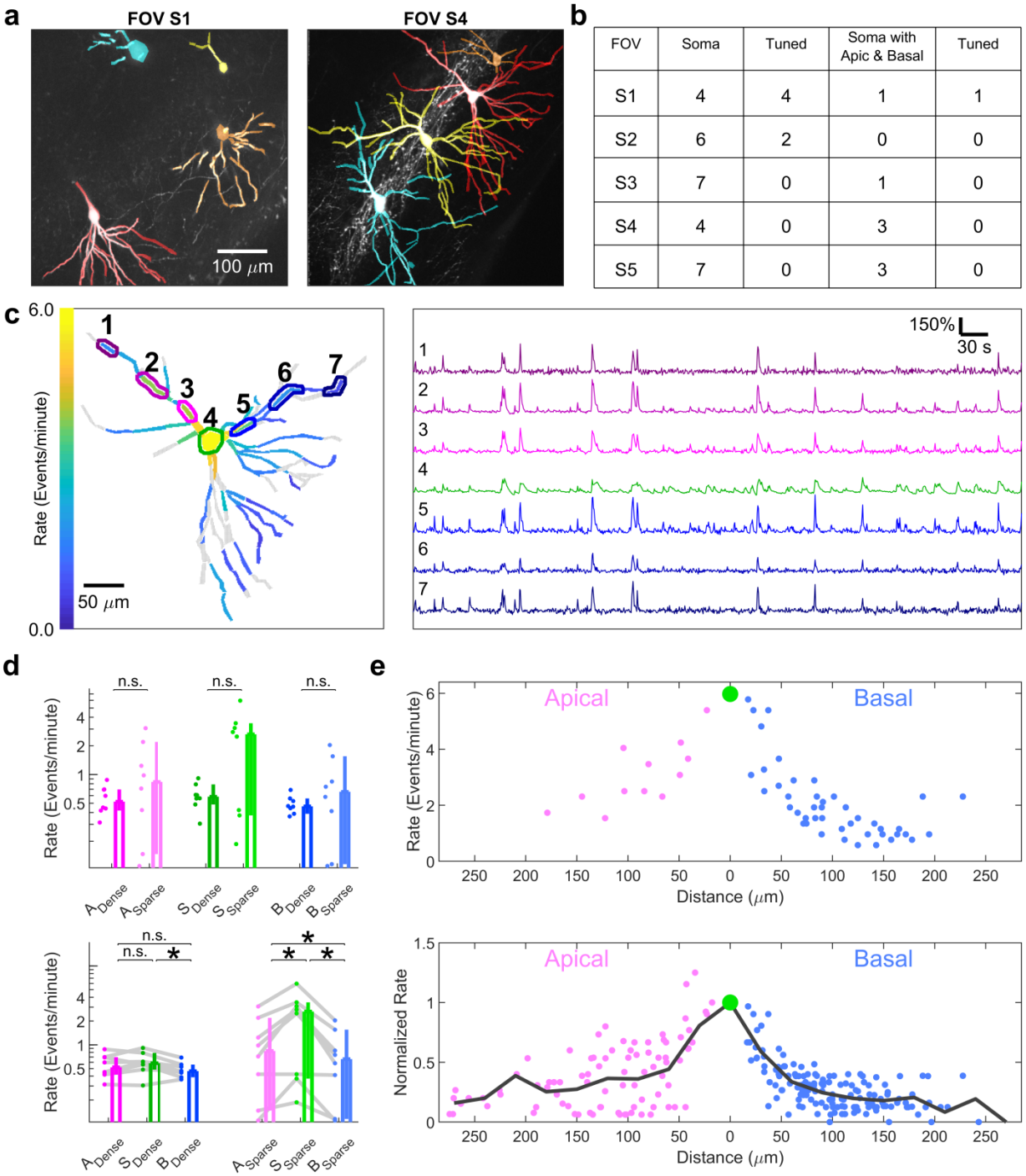
A8. The reviewer is right - confirming our results with sparse labeling is indeed a very important step. Taking this to heart, we have included a new data set from 5 fields of view from 4 mice, all with more sparsely labeled neurons (**Extended Data Fig. 12**). All told, we recorded 28 neurons across these fields of view, though only 8 had visible apical and basal dendritic trees (**Extended Data Fig. 12b-c**). In these neurons, we did not find any statistically significant differences between dense and sparse data in the apical, somatic, or basal compartments for event rate (**Extended Data Fig. 12d, Top**), validating the use of our sparse data. Moreover, we observe lower activity rates in the basal dendrites compared to the soma (**Extended Data Fig. 12d, Bottom**), supporting our original finding. In the sparse data, apical dendrites also had lower activity rates than the soma, but larger than the basal dendrites. These additional trends are visible in the dense data but do not reach statistical significance. Using our sparse data, we observed that the activity rate in both apical and basal dendrites decreases with distance from the soma (**Extended Data Fig. 12e**), which would naturally lead to lower activity rates in dendrites compared to soma.

We also attempted to compare spatial coding properties and stability between dense and sparse data. Calculating place field width or stability requires an ROI to be tuned, but in these sessions only 6 soma were tuned, and only 1 had visible apical and basal dendrites. For tuned ROIs, the place field width and stability were not statistically significantly different between the dense and sparse data (**Extended Data Fig. 13**). Basal dendrites tended to have smaller place field width compared to soma in the sparse data, but this was not statistically significant, as in the dense data. Apical dendrites tended to have more across-session stability compared to basal dendrites in the sparse data, but this also was not statistically significant. This underscores the difficulty in collecting enough data to arrive at statistically robust results using sparsely labeled data. Indeed, this is one of the motivations we have laid out for using densely labeled data, which emphasizes the need for our method. Because of the low yield, we feel like these analyses would bring limited additional value to the manuscript.

From the revised manuscript:

"To provide support that these findings are not an artifact of the dense recording preparation or our analytical approaches, we analyzed a subset of the data with sparse GCaMP expression. We found no differences in overall activity rates (Extended Data Fig. 12) or place field widths (Extended Data Fig. 13) between dense and sparse datasets in the apical, somatic, or basal compartments. Basal dendrites in the sparse data had lower rates than the soma, as in the dense data. In the sparse data, apical dendrites also had lower activity rates than the soma, but larger than basal dendrites. These additional trends were visible in the dense data but did not reach statistical significance. In the sparse data we observed that the activity rate in both apical and basal dendrites decreased with distance from the soma, which would naturally lead to lower activity rates in dendrites compared to soma. The similar findings in dense and sparse data validate that we can accurately detect calcium transients in the dense data."

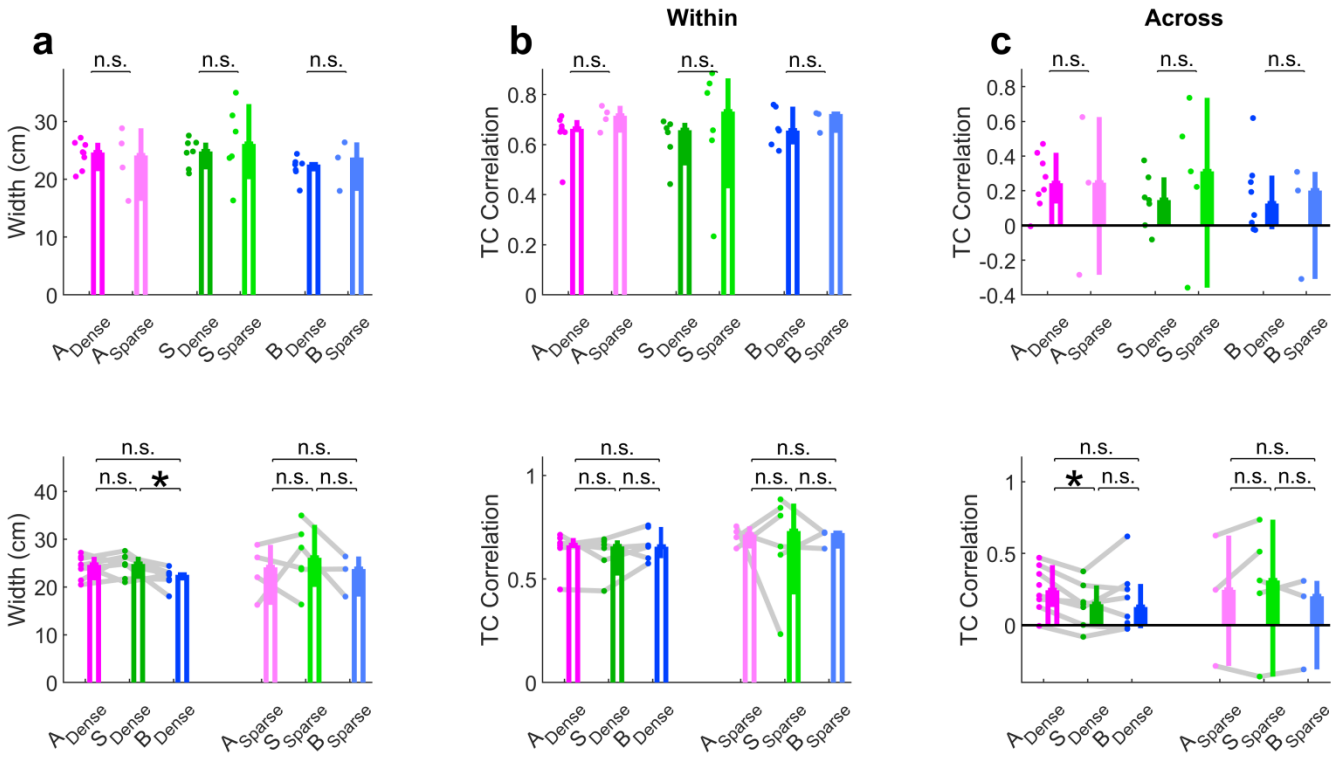
"We investigated stability properties in the sparse data as well, but the small amount of tuned neurons limited our ability to observe statistically significant differences (Extended Data Fig. 13)."



FOV	Soma	Tuned	Soma with Apic & Basal	Tuned
S1	4	4	1	1
S2	6	2	0	0
S3	7	0	1	0
S4	4	0	3	0
S5	7	0	3	0

Extended Data Fig. 12 | Sparsely labeled data reproduces finding of lower rates in basal dendrites. Maximum intensity projections of two sample fields of view from sparsely labeled mice, with neuron skeletons overlaid in different colors. **b.** Table showing the number of soma obtained in each field of view. **c.** Left, sample neuron (red highlighted neuron from FOV S4) pseudocolored with the event rate in each compartment. 7 sample ROIs are indicated with outlines. Right, fluorescence traces for the 7 indicated ROIs. Note there are more calcium transients the closer the ROI is to the soma (ROI 4, green). Activity is plotted as ΔS/S, where S is the ratio of green GCaMP to red tdTomato (see Methods). **d.** Top, overall activity rates were comparable between dense and sparse data for apical dendrites (Dense: 0.52, [0.42, 0.70]; Sparse: 0.84, [0.15, 2.2], p=0.38, Wilcoxon rank-sum test), soma (Dense: 0.59, [0.49, 0.79]; Sparse: 2.65, [0.37, 3.46], p=0.33, Wilcoxon rank-sum test), and basal dendrites (Dense: 0.46, [0.40, 0.56]; Sparse: 0.66, [0.11, 1.56], p=0.38, Wilcoxon rank-sum test). Bottom, basal dendrites had significantly lower rates than soma in the dense data. This was observed in the sparse data as well (Basal vs Soma: $p=7.8\times10^{-3}$, Wilcoxon sign-rank test), with the additional observation of apical dendrites having higher rates than basal dendrites but lower than the soma (Apical vs Basal: p=0.016, Wilcoxon sign-rank test; Apical vs Soma:

p=0.016, Wilcoxon sign-rank test). **e.** Top, activity rate as a function of distance from the soma for the example neuron in **c.** Bottom, aggregated normalized rates for all recorded neurons in the sparse datasets. Black line indicates the average in 30 μm bins.



Extended Data Fig. 13 | Place field width and stability in the sparse data. a. Top, place field width was not significantly different between dense and sparse datasets (Apical: Dense: 24.6, [22.1, 26.3] cm; Sparse: 24.1, [16.3, 28.8] cm; p=1, Wilcoxon rank-sum test; Soma: Dense: 24.8, [21.7, 26.3] cm; Sparse: 26.1, [20.0, 33.0] cm; p=0.63; Basal: Dense: 22.5, [21.4, 23.0] cm; Sparse: 23.7, [18.0, 36.4] cm; p=0.67), but there were limited tuned ROIs in the sparse data (4 apical dendritic trees, 6 soma, 3 basal dendritic trees). Bottom, basal dendrite place field width was smaller than somatic place field width in the dense data (p=0.047). This trend was present but not statistically significant in the sparse data (Soma vs Basal: p=0.5; Apical vs Soma: p=1; Apical vs Basal: p=1). Dense data are reproduced from Figure 6g. **b.** As in **a** but for within-session tuning curve correlation. No significant differences were present between Dense and Sparse data for any compartment (Apical: Dense: 0.66, [0.65, 0.70]; Sparse: 0.71, [0.65, 0.75]; p=0.23; Soma: Dense: 0.66, [0.52 0.69]; Sparse: 0.73, [0.43, 0.86]; p=0.48; Basal: Dense: 0.66, [0.60, 0.75]; Sparse: 0.72, [0.65, 0.73]; p=0.83), and no significant differences were present between compartments for either dataset (Dense: Apical vs Soma: p=1; Apical vs Basal: p=1; Soma vs Basal: p=0.31; Sparse: Apical vs Soma: p=1; Apical vs Basal: p=1; Soma vs Basal: p=0.75). Dense data are reproduced from Figure 7f. **c.** As in **a** but for across-session tuning curve correlation. No significant differences were present between Dense and Sparse data for any compartment (Apical: Dense: 0.24, [0.13, 0.42]; Sparse: 0.25, [-0.28, 0.63], N=3 dendritic trees; p=1; Soma: Dense: 0.15, [0.00 0.28]; Sparse: 0.31, [-0.36, 0.74], N=5 soma; p=0.27; Basal: Dense: 0.13, [-0.02, 0.29]; Sparse: 0.20, [-0.31, 0.31], N=5 dendritic trees; p=1). In the Dense data, apical dendrites were significantly higher than soma (p=0.016), but this was not the case in the Sparse data (Apical vs Soma: p=0.5; Apical vs Basal: p=1; Soma vs Basal: p=1). Dense data are reproduced from Figure 7f.

Q9. *Finally, the manuscript contains minor errors that would benefit from some attention. Some figure panels are missing, and some figure legends do not accurately reflect the figures themselves.*
There are instances where terms and acronyms are used before being accurately defined.
In addition, some sentences could be phrased more clearly, such as references to "the area under the curve" without specifying which curve, and statements about "high skewness values" without clarifying the corresponding distribution. Addressing these small issues would make the manuscript easier to read.

A9. We thank the reviewer for pointing out these inconsistencies and apologize for the oversights. Figures 2 and 3 (now Figures 2 and 4) had correct figure legends but not up-to-date figures, and references to Figure 1 were outdated. These are now fixed, and the usage of terms such as "area under the curve" and "high skewness values" has been clarified.

Reviewer #2

Q10. *In this paper, Moore et al describe a modification of the CNMF algorithm for finding more dendritic ROIs in population calcium imaging of neuronal compartments.*

This paper does not address the main problem of dendritic imaging, which is the backpropagating action potential, and thus I have difficulty understanding who the method is targeted towards.

Studies of dendritic activity are always done with respect to the somatic activity, because the backpropagating action potential typically swamps out the smaller and more rare dendritic spikes. Identifying dendritic spikes that are not just backpropagating APs is thus the main problem of dendritic imaging, and this must be resolved if the present method is to be useful.

A10. The dissociation of bAPs in dendritic activity is indeed an important problem for the field. We acknowledge that we did not explicitly address this issue, but feel that these problems are outside the scope of our study. Here we tackle the issue of accurately estimating the time-varying calcium activity in preparations densely packed with overlapping populations of neurons and their processes. Dendrites provide an excellent model for densely labeled preparations, as the issues arising from overlapping ROIs are likely to be more apparent in thin processes compared to large soma.

The scientific findings we report comparing the basic activity profiles and spatial tuning features of apical and basal dendritic populations do not rely on resolving bAPs from local sources. Untangling and identifying the sources of dendritic time-varying activity (backpropagating versus local) is best addressed via experimental manipulations with known anatomical ground truth connecting dendrites and soma. Further, explicit, concrete identification of forward- versus back-propagating signals would likely require voltage indicators given rapid propagation speeds.

Q11. *In many cell types, dendritic spikes are not even present, or may be exceedingly rare. Are the authors convinced that dendritic spikes exist in these particular recordings?*

A11. Based on patch-clamp recordings from CA3 (Kim, Guzman, Hu, Jonas 2012; Moore, Robert, Rashid, Basu 2022), we are confident that local dendritic events that do not propagate to the soma do exist. However, as noted in the above points, their presence or absence is somewhat orthogonal to the utility of this approach.

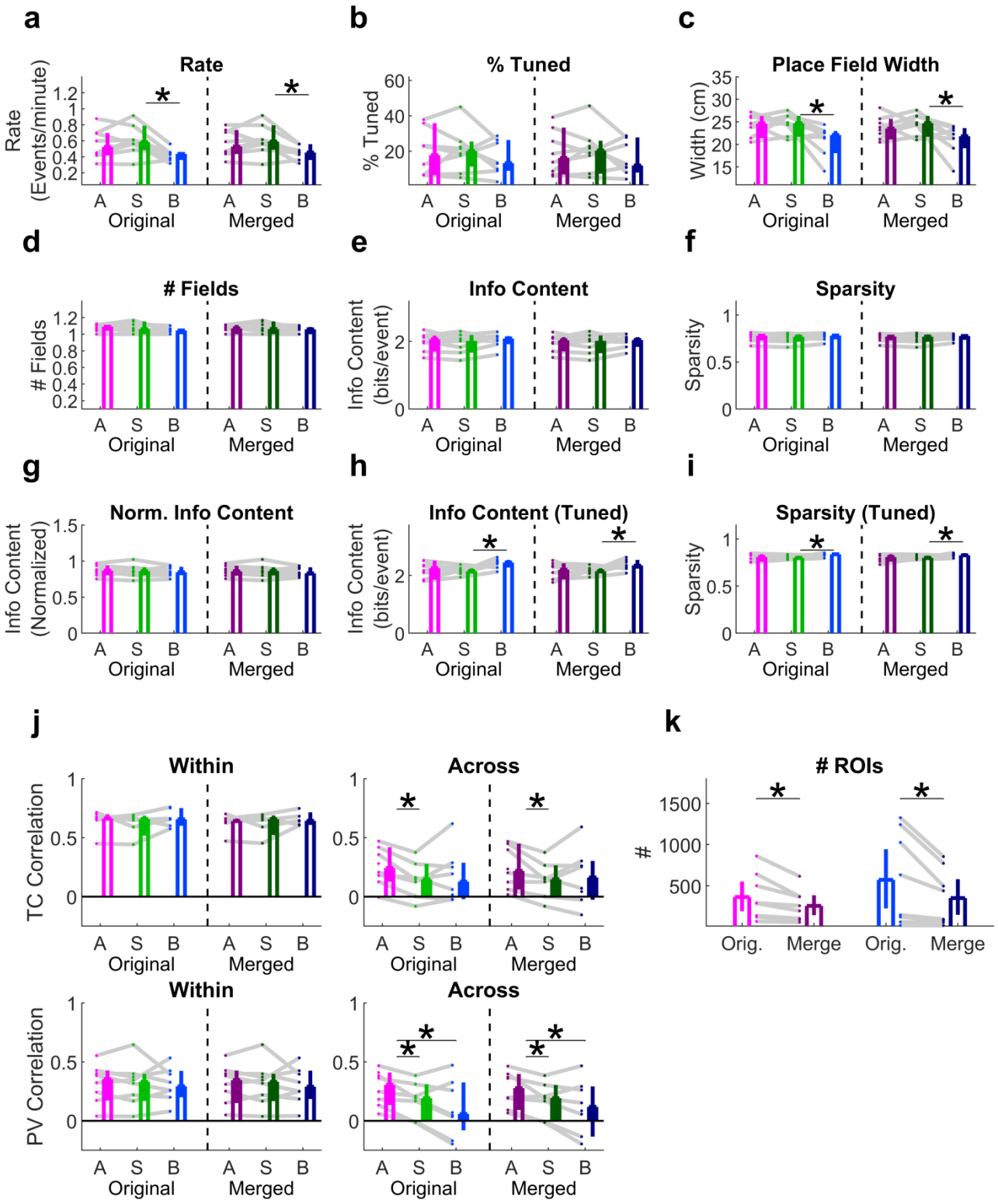
Q12. *The authors also do not address the oversplitting problem, where multiple dendrites from the same neuron are split into distinct ROIs. Combining these ROIs is necessary for follow-up analyses, since otherwise the neurons with many oversplit dendrites are counted more times than the neurons with undersplit dendrites. An automated method to do this would be welcome, and benchmarks should be done with respect to fully connected neurons, not individual dendrite pieces.*

A12. This is indeed an important factor to keep in mind when performing statistical comparisons. However, depending on the experimental question, oversplitting can be more or less an issue that can be addressed with statistical approaches such as sub-sampling.

To explore if this affected our scientific results, we re-computed all population measures by merging ROIs of the same type (apical, soma, basal) that had an activity correlation above 0.8 (**Extended Data Fig. 11**). This cut the number of apical dendrites by an average of 30% and the number of basal dendrites by an average of 40%. All statistical comparisons were identical in the original and merged datasets, indicating that oversplitting did not affect our results. It may be interesting to experiment with different thresholds for merging, but we believe this is beyond the scope of this study.

From the revised manuscript:

"Furthermore, these results were not based by "oversplitting" of dendrites into multiple compartments, as all comparisons were similar when we merged ROIs with highly correlated activity (see Methods, Extended Data Fig. 11)."



Extended Data Fig. 11. Oversplitting does not affect statistical measures. a-j, Statistical measures show similar trends and significant differences between the original data and data obtained from merging ROIs with similar activity. See Table 5 for summary statistics. **k,** The number of apical dendrites in the merged dataset (260, [140, 380], n=8 FOVs) was ~30% smaller than the number in the original dataset (360, [190, 550]). Similarly, the number of basal dendrites in the merged dataset (350, [150, 580]) was ~40% smaller than the number in the original dataset (570, [220, 950]).

Panel	Measure	Apical	Soma	Basal	A vs S	A vs B	S vs B
a	Rate (Orig.)	0.52, [0.42, 0.70]	0.59, [0.49, 0.79]	0.44, [0.36, 0.46]	P=0.38	p=0.078	p=0.023
	Rate (Merge)	0.53, [0.44, 0.74]	0.59, [0.49, 0.79]	0.45, [0.36, 0.56]	p=0.38	p=0.11	p=0.023
b	% Tuned (Orig)	17.8, [6.6, 35.8]	20.1, [7.3, 25.5]	13.2, [9.0, 26.3]	p=0.84	p=0.74	p=0.31
	% Tuned (Merge)	16.2, [6.9, 33.3]	20.4, [7.3, 25.9]	11.9, [8.0, 27.8]	p=0.74	p=0.64	p=0.31
c	Place Field Width (Orig.)	24.6, [21.4, 26.3]	24.7, [21.7, 26.3]	22.1, [18.0, 23.0]	p=1	p=0.078	p=0.023
	Place Field Width (Merge)	23.5, [21.1, 25.7]	24.7, [21.7, 26.3]	21.8, [19.1, 23.6]	p=0.55	p=0.15	p=0.023
d	# Fields (Orig.)	1.09, [1.04, 1.11]	1.06, [1.00, 1.15]	1.05, [1.00, 1.07]	p=0.94	p=0.25	p=0.3
	# Fields (Merge)	1.07, [1.00, 1.11]	1.06, [1.01, 1.14]	1.06, [1.00, 1.09]	p=0.69	p=0.81	p=0.38
e	Info Content – All (Orig.)	2.06, [1.74, 2.17]	2.01, [1.67, 2.19]	2.09, [1.92, 2.16]	p=0.31	p=0.46	p=0.15
	Info Content – All (Merge)	2.03, [1.71, 2.13]	2.02, [1.65, 2.18]	2.05, [1.84, 2.13]	p=0.38	p=0.46	p=0.25
f	Sparsity – All (Orig.)	0.78, [0.73, 0.80]	0.77, [0.72, 0.80]	0.78, [0.74, 0.80]	p=0.15	p=0.55	p=0.25
	Sparsity – All (Merge)	0.77, [0.73, 0.79]	0.77, [0.72, 0.79]	0.78, [0.74, 0.79]	p=0.38	p=0.55	p=0.31
g	Norm. Info Content – All (Orig.)	0.87, [0.80, 0.94]	0.86, [0.80, 0.91]	0.84, [0.80, 0.92]	p=1	p=0.55	p=0.55
	Norm. Info Content – All (Merge)	0.86, [0.80, 0.93]	0.86, [0.80, 0.91]	0.83, [0.80, 0.91]	p=0.64	p=0.55	p=0.46
h	Info Content – Tuned (Orig.)	2.25, [1.85, 2.51]	2.15, [2.07, 2.25]	2.45, [2.27, 2.54]	p=0.46	p=0.11	p=0.016
	Info Content – Tuned (Merge)	2.20, [1.86, 2.42]	2.14, [2.07, 2.25]	2.35, [2.24, 2.54]	p=0.74	p=0.2	p=0.039
I	Sparsity – Tuned (Orig.)	0.81, [0.76, 0.84]	0.80, [0.79, 0.82]	0.84, [0.83, 0.85]	p=0.95	p=0.078	p=0.016
	Sparsity – Tuned (Merge)	0.81, [0.76, 0.84]	0.80, [0.79, 0.82]	0.83, [0.82, 0.85]	p=0.64	p=0.15	p=0.039
j	TC Correlation – Within (Orig.)	0.66, [0.65, 0.70]	0.66, [0.52, 0.69]	0.66, [0.60, 0.75]	p=1	p=1	p=0.31
	TC Correlation – Within (Merge)	0.64, [0.62, 0.66]	0.66, [0.52, 0.69]	0.64, [0.61, 0.71]	p=0.69	p=0.56	p=0.44
	TC Correlation – Across (Orig.)	0.24, [0.13, 0.42]	0.15, [0.00, 0.28]	0.13, [-0.02, 0.29]	p=0.016	p=0.11	p=0.38
	TC Correlation – Across (Merge)	0.22, [0.06, 0.45]	0.15, [0.00, 0.27]	0.16, [-0.03, 0.30]	p=0.031	p=0.15	p=0.58
	PV Correlation – Within (Orig.)	0.35, [0.17, 0.43]	0.33, [0.17, 0.40]	0.29, [0.20, 0.42]	p=0.74	p=0.95	p=1

	PV Correlation – Within (Merge)	0.35, [0.15, 0.43]	0.33, [0.17, 0.40]	0.29, [0.18, 0.42]	p=0.95	p=0.46	p=0.95
	PV Correlation – Across (Orig.)	0.30, [0.13, 0.41]	0.19, [-0.01, 0.31]	0.06, [-0.17, 0.33]	p=0.016	p=0.039	p=0.16
	PV Correlation – Across (Merge)	0.28, [0.09, 0.40]	0.19, [-0.01, 0.30]	0.12, [-0.07, 0.29]	p=0.031	p=0.039	p=0.3
Panel	**Measure**	**Original**		**Merged**		**Comparison**	
k	# ROIs (Apical)	360, [190, 550]		260, [140, 380]		p=0.0078	
	# ROIs (Basal)	570, [220, 950]		350, [150, 580]		p=0.0078	

Table 5 | Statistics for Extended Data Fig. 11.

Q13. *What is shown instead in this paper is an improvement over previous methods in terms of the raw quantity of ROIs that are detected, which is good but not particularly exciting.*

The failure to address the backpropagating AP and the oversplitting of the dendrites is a problem for all existing algorithms, and something they all fail at. However, I would expect these problems to be resolved by a method that claims to be specially developed for dendrites.

A13. The problems we set out to solve were automatic detection of dendritic ROIs and accurate estimate of their time-varying calcium activity, as discussed above. Notably, other methods fail at this, as noted by Reviewer 3, so this is a valuable contribution to the field.

Reviewer #3

Q14. *In this manuscript, the authors present an algorithm, d-NMF, to automatically detect somatic and dendritic regions of interest (ROIs) in a field of view (FOV), in which neuronal processes are densely labeled with a genetically encoded Ca2+ indicator (either GCaMP6 or 7b). Two-photon Ca2+ imaging datasets at varying labeling densities, obtained from hippocampal area CA3, an area challenging to access by imaging, are used to highlight the algorithm's properties. The nature of this specific preparation (pyramidal neurons in area CA3 (probably subarea a) can be imaged sidewise) allows tracking of large populations of dendrites from different compartments. Notably, this would not be the case for other brain areas.*

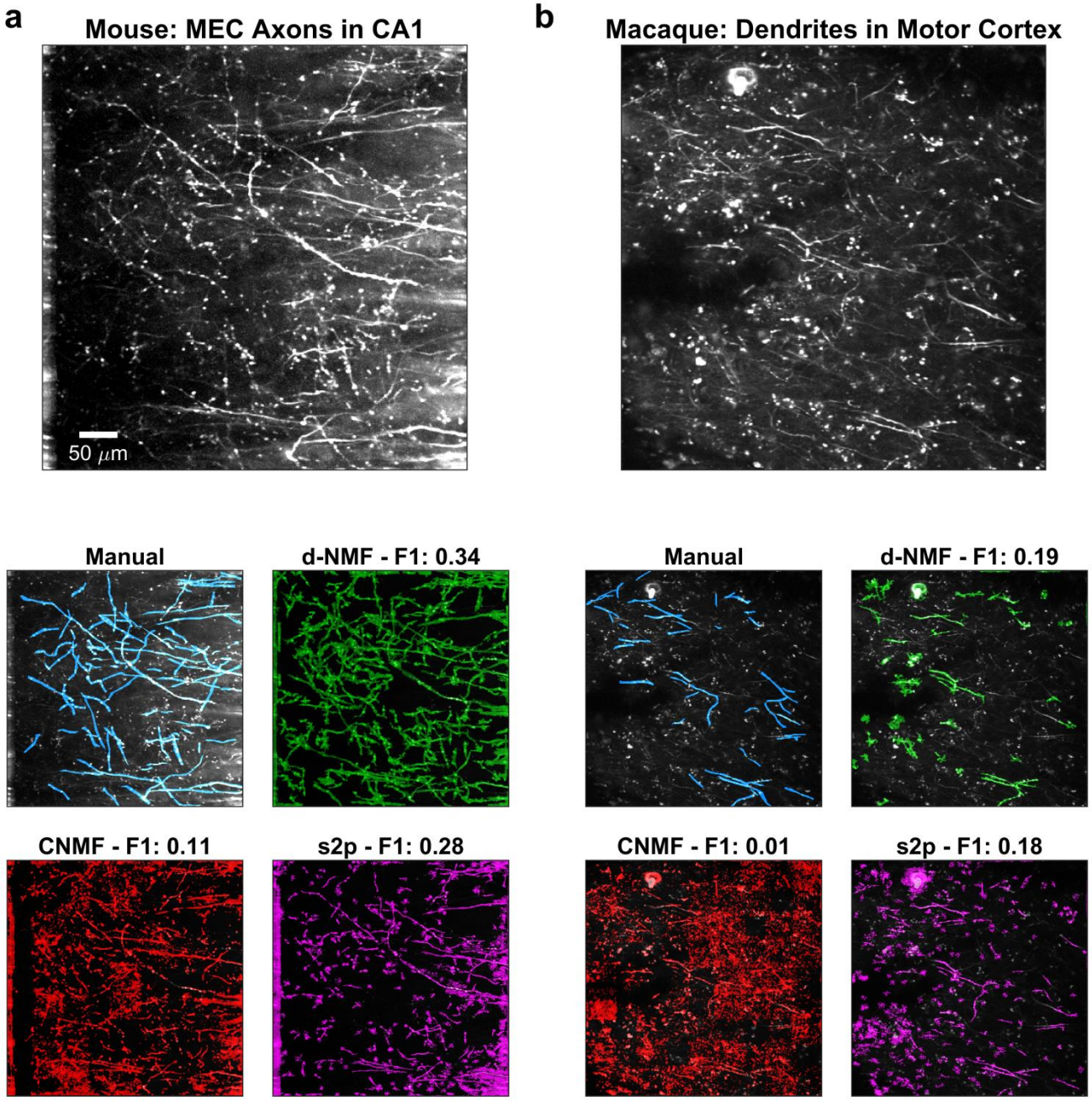
A14. Many cortical regions have the same geometrical affordance, such as retrosplenial cortex. Furthermore, microprism approaches can be used to achieve similar fields of view. Finally, tuft dendrite imaging or even basal dendrite imaging could benefit from our approaches.

We have now included two other datasets, one from Medial Entorhinal Cortex axons in CA1 of the mouse hippocampus from our lab, and the other a publicly available dataset from the motor cortex of a macaque monkey (Trautmann et al., Nature Communications 2021). These are difficult datasets to process, but as shown in the attached **Extended Data Fig. 5** we outperform other methods with respect to ROI accuracy in these diverse datasets.

From the revised manuscript:

"d-NMF also matched or outperformed other methods when applied to other types of data, including dendrites from motor cortex in macaque monkeys and entorhinal cortex axons in mouse CA1 (Extended Data Fig. 5)."

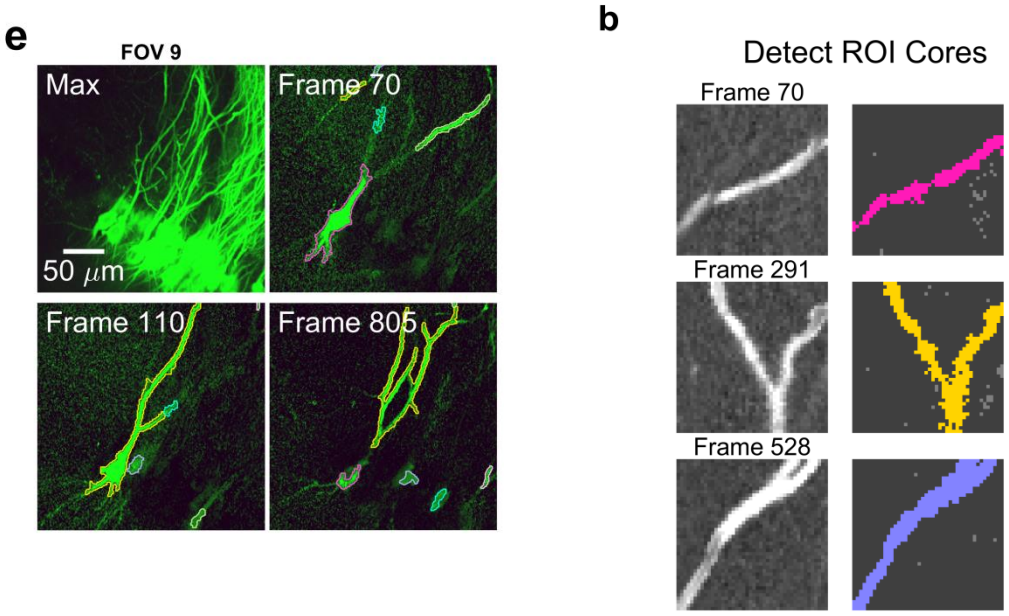
"Hippocampal area CA3 provides the geometrical advantage of imaging large dendritic trees in a single focal plane. Our approach would easily extend to brain areas such as retrosplenial cortex with similar affordances, or those accessible with microprism approaches. Our method can be utilized in more traditional tuft dendrite or basal dendrite imaging, and could facilitate analysis of sparsely labeled preparations as well."



Extended Data Fig. 5. Application to other datasets. a, Sample field of view with axons from Medial Entorhinal Cortex (MEC) labeled with GCaMP6s, imaged in the Stratum Lacunosum Moleculare of mouse CA1. Top represents the maximum projection image. The panels in the bottom are as in Figure 3, illustrating ROIs obtained from manual labeling, d-NMF, CNMF, and suite2p. **b**, Sample field of view with dendrites in macaque motor cortex labeled with GCaMP6f.

Q15. *The algorithm's results are validated by comparison with human expert manual labeling. However, it is doubtful, in my opinion, that dense FOVs such as those shown in Extended Data Fig. 1 (FOVs 11-14) can be segmented accurately.*

A15. An important aspect of our dense datasets is that they are sparsely active in time, such that very few ROIs are active at any given time. This enables an experimenter (or algorithm) to accurately segment dense fields of view by going through the imaging t-series movie frame by frame. This point is explored in **Figures 1e and 2b** (copied below), but we emphasize this more in the revised manuscript. We also include a new supplemental movie illustrating the point that though the dendrites may be dense in space they are relatively sparse in time. As a result, only a small number of neurons show elevated activity on any given frame, which allows accurate segmentation.



Left, Figure 1e, demonstrating that in a dense FOV (maximum projection, top left), few ROIs are active on individual frames. Right, Figure 2b, additional examples of overlapping dendrites being active only on particular frames.

From the revised manuscript:

"Despite the highly overlapping nature of dendrites in the more densely labeled samples, manual segmentation was possible because the neurons were sparsely active in time, such that only a few neurons were active on any given frame (Fig. 1e, Supplementary Movie 1), reducing the dense segmentation problem into a series of sparse segmentation problems."

Q16. *Dendritic Ca2+ imaging is commonly used to understand how the dendritic processing of synaptic inputs shapes an individual neuron's action potential (AP) output. Usually, these experiments are performed in sparsely labeled FOVs because this allows 1) tracking of the dendritic ROIs over time and 2) assigning dendritic branches to a soma unambiguously. While the paper addresses the first point appropriately, the second, and probably more essential, the issue is not addressed by d-NMF (as briefly discussed in the 'discussion' section).*

A16. We agree that unambiguously assigning dendritic branches to soma is an important goal, but one that we did not try to solve here. One issue is that many dendrites in a field of view may have a parent soma in a different focal plane. Furthermore, with just a functional indicator this may not be possible to solve to a high degree of rigor. Future experiments with an additional structural marker can address this.

From the revised manuscript:

"Additional structural markers would likely help in linking connected dendrites and soma. Furthermore, many distal dendrites are connected to soma that are not in the focal plane, which could only be resolved with multi-planar imaging. These issues of connecting dendrites with parent soma are beyond the scope of this study, but will be critical problems to address in the future."

Q17. *I am generally not convinced that imaging FOVs with densely labeled dendrites can provide valuable insights into dendritic integration and its role in single-neuron function. However, I see the benefits of using d-NMF in automatically segmented sparsely labeled dendritic fields (often, algorithms focused on somas fall short here). I would suggest emphasizing this aspect much more in the paper.*

A17. It is indeed difficult to connect all dendrites to their parent soma, so we agree that the method as presented is better suited to automatically detecting dendritic ROIs rather than investigating single-neuron functions. However, this approach does provide information about dendritic tuning in general, which is a necessary first step to investigating dendrite-soma interactions.

From the revised manuscript:

"d-NMF performed well on both sparse and dense datasets with a single set of parameters, whereas other methods may need careful parameter tuning depending on the density of the preparation."

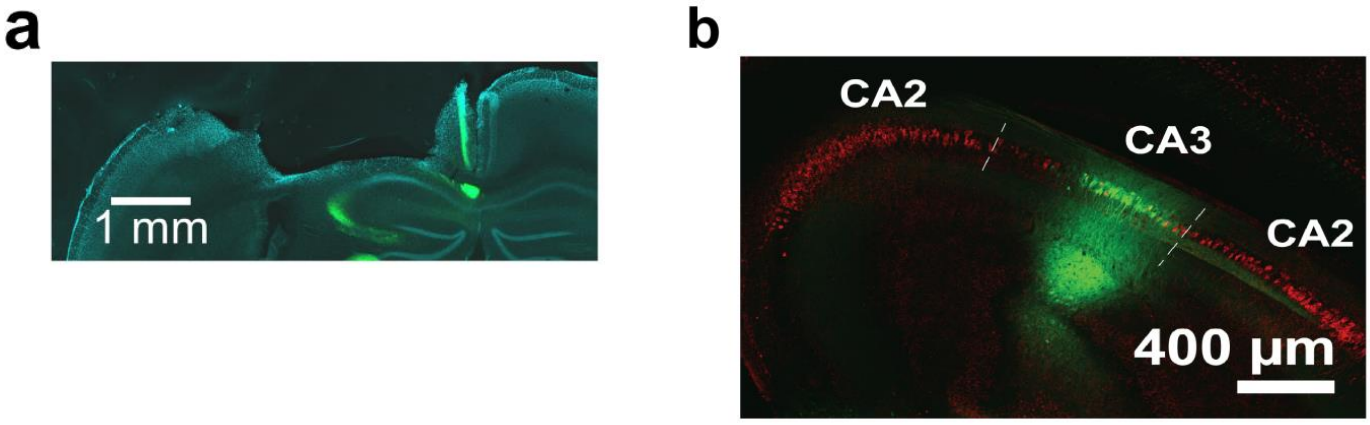
"Our method can be utilized in more traditional tuft dendrite or basal dendrite imaging, and could facilitate analysis of sparsely labeled preparations as well. The high accuracy of d-NMF in sparsely labeled datasets will allow automatic extraction and correlation of dendritic and somatic activity in an unbiased manner. This will facilitate investigation of dendritic integration, soma-dendrite coupling dynamics, signal propagation, and their roles in single-neuron function."

"Our approach already merges together ROIs of neighboring dendrites and soma which have substantial spatial overlap and highly correlated activity. Such ROIs obtained from merging the soma and dendrites indicate a high degree of coupling between the soma and its proximal dendritic branches, as would be expected from the diffusion of calcium over short distances."

Q18. *In addition, evidence and discussion should be added to address the following issues:*
1) Throughout the paper, critical basic experimental descriptors need to be included. For example, what is the path distance from the soma to the given imaged dendritic ROI? Is there a systematic difference in path distance between basal and apical dendrites imaged? Which region of CA3 was imaged?

A18. With regards to the path distance from the soma and dendrites, this is not a straightforward question. Each FOV contains many dendrites and soma, and any summary value of a session is an average over all ROIs within that FOV. We did quantify the distance of ROIs from the estimated cell body layer. Apical dendrite ROIs had a mean distance 110 um from the cell body layer, with a maximum of 300 um. Basal dendrite ROIs had a mean distance of 70 um from the cell body layer, with a maximum of 160 um. However, we do not explicitly link all dendrites to their parent soma, the parent soma of any given dendrite is not always visible. We can include these numbers in the manuscript, but we do not feel like these crude distance measures provide critical information.

Based on our injection coordinates (1.5 mm lateral, 1.3 mm caudal of Bregma; 1.7 mm lateral, 1.5 mm caudal of Bregma; depths of 1.8 and 2 mm below the dural surface), our field of views with the orientation of the neurons and their dendrites, as well as immunohistochemistry for PCP4, a stain specific to CA2 (**Extended Data Fig. 1b**), we can confirm we imaged in CA3a/b. Please also note the location of GCaMP6f in green in the histology figure in **Fig. 1a**.



Left, Figure 1a, showing GCaMP6f in green in area CA3a/b. Right, Extended Data Figure 1b, showing a representative horizontal slice showing GCaMP6f (green) is expressed in area CA3, not CA3, marked by PCP4 (red).

[REDACTED]

Images from the Paxinos mouse brain atlas (“The Mouse Brain in Stereotaxic Coordinates, Third Edition”). Approximate injection coordinates are marked with red circles.

Q19. *2) How are ROIs defined? As far as I can tell, ROI detected with d-NMF can contain multiple branches, single branches, and areas that contain only part of a branch. That means that the nature of these Ca2+ signals will be a complete mix: it could be significant single-spine signals, local dendritic-spike associated signals, or it could be backpropagating APs. Can d-NMF address this problem?*

A19a. Automatically detected ROIs can in theory contain multiple branches, but any identified ROI is designed to have synchronous activity across all of its pixels. So any branches that are actually showing independent activity will be detected as separate ROIs. See our response to point Q6 above and Q22 below.

From the revised manuscript:

"The resulting ROIs also have the property that their activity is homogenous within an ROI. Any dendritic branch or portion of a large ROI that has independent activity is split off into its own ROI (Extended Data Fig. 3)."

A19b. The nature of the calcium signals, local or backpropagating, is not addressed here and may not be addressable with calcium imaging. We provide an estimate of the time course of calcium. Any further interpretation is up to the experimenter. Experiments can be designed to test this, and d-NMF can be used as a signal extraction tool. Also see our responses to points Q10 and Q13.

Q20. *3) Using skewness as a parameter to classify ROIs introduces a clear bias in ROI detection. It will also limit the cell types for which this algorithm can be used.*

A20. This is true. Our threshold was optimized for CA3 pyramidal neurons, and it would need to be adjusted for different regions or cell types. However, we did also use the Signal-to-Noise ratio to classify ROIs, with similar results.

From the revised manuscript:

"Similar results were obtained using the signal-to-noise ratio (SNR) for classification (Extended Data Fig. 4)."

Q21. *4) Along the same lines, the findings related to event rates and spatial selectivity in the different dendritic compartments need to be validated via imaging in sparsely labeled Ca2+ imaging datasets. For example, how can event rates in basal dendrites (a dendritic region where dendritic Ca2+ signals are known to be dominated by backpropagating action potentials) differ from somatic event rates, which are known to be driven by action potentials? Could this be a result of the variable ROI sizes and patterns?*

A21a. The reviewer is right - confirming our results with sparse labeling is indeed a very important step. Taking this to heart, we have included new data from 5 fields of view from 4 mice, all with more sparsely labeled neurons (**Extended Data Fig. 12a**). All told, we recorded 28 neurons across these fields of view, though only 8 had visible apical and basal dendritic trees (**Extended Data Fig. 12b-c**). In these neurons, we did not find any statistically significant differences between dense and sparse data in the apical, somatic, or basal compartments (**Extended Data Fig. 12d, Top**), validating the use of our sparse data. We do observe lower activity rates in the basal dendrites compared to the soma (**Extended Data Fig. 12d, Bottom**), supporting our original finding. In the sparse data, apical dendrites also had lower activity rates than the soma, but larger than the basal dendrites. These additional trends are visible in the dense data but do not reach statistical significance. Using our sparse data, we observed that the activity rate in both apical and basal dendrites decreases with distance from the soma (**Extended Data Fig. 12e**), which would naturally lead to lower activity rates in dendrites compared to soma.

We also attempted to compare spatial coding properties and stability between dense and sparse data. Calculating place field width or stability requires an ROI to be tuned, but in these sessions only 6 soma were tuned, and only 1 had visible apical and basal dendrites. For tuned ROIs, the place field width and stability were not statistically significantly different between the dense and sparse data (**Extended Data Fig. 13**). Basal dendrites tended to have smaller place field width compared to soma in the sparse data, but this was not statistically significant, as in the dense data. Apical dendrites tended to have more across-session stability compared to basal dendrites in the sparse data, but this also was not statistically significant. This underscores the difficulty in collecting enough data to arrive at statistically robust results using sparsely labeled data. Indeed, this is one of the motivations we have laid out for using densely labeled data, which underscores the need for our method. Because of the low yield, we feel like these analyses would bring limited additional value to the manuscript.

Figures 12 and 13 are reproduced above, in response to Q8.

From the revised manuscript:

"To provide support that these findings are not an artifact of the dense recording preparation or our analytical approaches, we analyzed a subset of the data with sparse GCaMP expression. We found no differences in overall activity rates (Extended Data Fig. 12) or place field widths (Extended Data Fig. 13) between dense and sparse datasets in the apical, somatic, or basal compartments. Basal dendrites in the sparse data had lower rates than the soma, as in the dense data. In the sparse data, apical dendrites also had lower activity rates than the soma, but larger than basal dendrites. These additional trends were visible in the dense data but did not reach statistical significance. In the sparse data we observed that the activity rate in both apical and basal dendrites decreased with distance from the soma, which would naturally lead to lower activity rates in dendrites compared to soma. The similar findings in dense and sparse data validate that we can accurately detect calcium transients in the dense data."

"We investigated stability properties in the sparse data as well, but the small amount of tuned neurons limited our ability to observe statistically significant differences (Extended Data Fig. 13)."

A21b. The control analysis we present in Extended Data Figure 9 suggests that ROI size is not a driving factor of the differences in rate.

Q22. *5) The fitness trace is an appropriate parameter to quantify if all Ca2+ events in a given ROI involved the entire ROI. But this is not true. What about, for example, single-branch signals or NMDA spike-mediated Ca2+ signals? This method would reject them because not the entire ROI is involved. This introduces bias, which should be at least discussed.*

A22. We believe this is a result of unclear communication on our part. It is true that the fitness curve approach requires the entire dendritic ROI to be activated for each event. However, this does not lead to a biasing of detection which would exclude fast localized events. This is because our algorithm defines ROIs as contiguous pixels that have synchronous activity. If one branch of a dendrite exhibits independent, localized activity, it will be treated as a separate ROI from its parent branch, and there will be no conflict.

We did not clearly articulate this point previously, but have taken steps to clarify it. We explore this using a new supplemental figure (**Extended Data Fig. 3**), which explores how the algorithm approaches 3 varying degrees of synchrony in a dendritic branch, from totally synchronous to heavily asynchronous. In all 3 cases the true activity of the branches was recovered.

Regarding the rate of detected events, it is absolutely true that the results can depend on the event detection criteria. **Main Figure 5**, which introduces the concept of the fitness trace, shows that using the fitness trace results in more accurate event identification than the 2z method (**Figure 5c**, attached here). We have expanded **Figure 5c** to include the actual even rate of manually-labeled events, and show that using the fitness trace results in an accurate estimate of the rate, and it is the 2Z method that (falsely) detects twice the number of events.

Extended Data Figure 3 and Figure 5c are reproduced above, in response to Q6

From the revised manuscript:

" The resulting ROIs also have the property that their activity is homogenous within an ROI. Any dendritic branch or portion of a large ROI that has independent activity is split off into its own ROI (Extended Data Fig. 3)."

"An implicit requirement of the Fitness Trace is that the entire ROI must be active for a calcium transient to be detected, which could potentially discard localized dendritic activity and bias estimates. If ROIs are too broadly defined, such that they span multiple branches containing independent activity, this would lead to a bias to exclude branch-independent transients. However, the d-NMF algorithm identifies ROIs precisely as groups of pixels with synchronous activity. Dendritic branches that show independent activity are automatically split off from other branches and analyzed as a separate ROI (Extended Data Figure 3). In this way accurate ROI identification assists in accurate transient detection."

Q23. *Minor points:*
1) Line 101: figure 1b does not have a single neuron labeled

A23. This was an oversight on our part and we have corrected this. Figure 1d has the single neuron labeled.

Q24. *2) Line 144: figure 2c does not exist; maybe this should be 3c?*

A24. The reviewer is correct that Figure 2c was misplaced as Figure 3c. We have updated the Figures 2 and 3 (now Figures 2 and 4) to be accurate to the text and figure legends

We have formatted this response as follows:

Original remarks from the reviewers in black italics

Our responses in blue

Quotations from the updated manuscript in red

Remarks and responses are numbered by Query (Q1, Q2, Q3…) and Answer (A1, A2, A3…)

For points common to more than one reviewer, we have addressed them with similar text and figures in each individual subsection.

Reviewer #1:

Q1. *We appreciate the great efforts the authors have made to address our concerns: many of the concerns have been addressed. In particular, the use of synthetic data to demonstrate how density affects the sensitivity of d-NMF and the use of a sparse preparation to validate their experimental findings are valuable additions to this study.*

A1. We thank the reviewer for their recognition of the work and for the helpful suggestions that have made the study stronger and our algorithm more versatile.

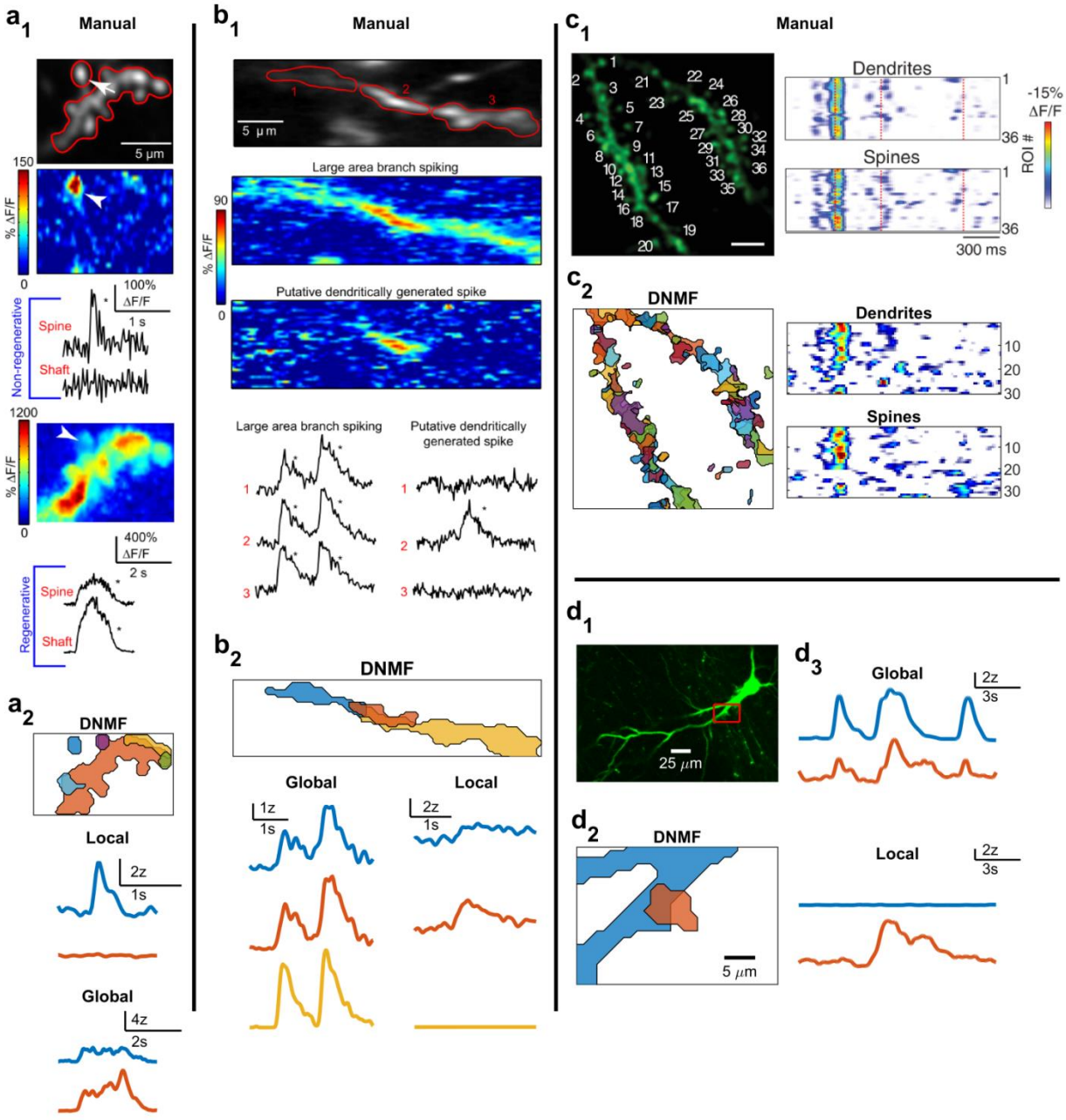
Q2. *However, I am still concerned about the ability of d-NMF to detect localized dendritic transients. Clarifying this point is critical for future users, as well as validating the experimental results presented in the manuscript. Low SNR and fast timescale localized dendritic events have previously been suggested to play a role in the hippocampus in memory formation.*

A2. We agree that the ability to detect localized dendritic transients in a wide variety of preparations is important for our approach to enjoy widespread use and success. We have taken this opportunity to demonstrate our method on 4 separate new datasets from 2 other pioneering labs that are experts in the field of dendritic two-photon imaging as well as experimental and synthetic data from our lab. Each data set shows instances of localized dendritic activity that our method detected and extracted.

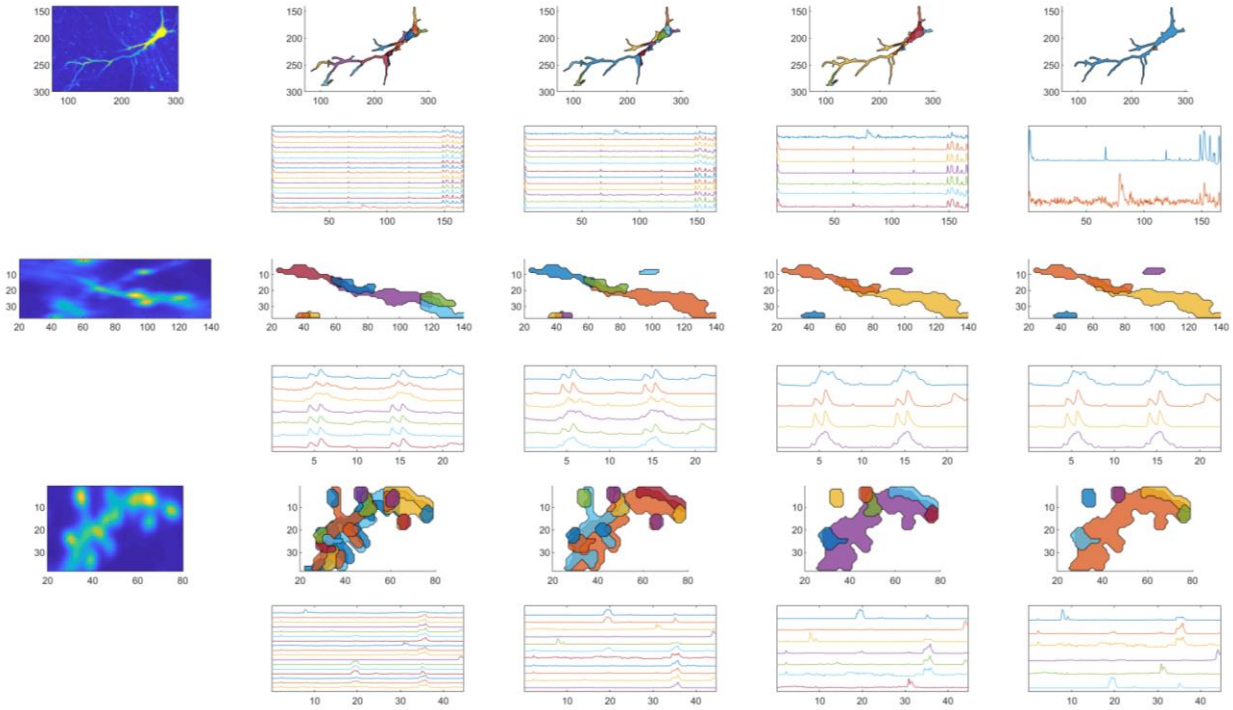
From the main text, lines 208-210: We validated that d-NMF was able to detect localized activity from two previously published datasets, using calcium imaging in mouse hippocampal area CA1[21] and voltage imaging in visual cortex[38], respectively. We also identified local activity in a putative dendritic spine in our own data (**Extended Data Fig. 7**).

To arrive at these ROIs, we adjusted the threshold at which to merge ROIs. This threshold can be adjusted by the used depending on how fine- or coarse-grained the user desires. This is illustrated in **Reviewer Fig. 1**.

Editorial Note: Content in panels a and b reproduced from https://doi.org/10.1038/nature13871, with permission from Springer Nature. Content in panel c reproduced from https://doi.org/10.1126/science.abg0501 with permission from AAAS.



Extended Data Fig. 7 | Fully automatic Identification of localized dendritic activity by d-NMF. a_1, From Sheffield & Dombeck, *Nature* 2015, illustration of a stretch of dendrite of a mouse CA1 pyramidal cell expressing GCaMP6f showing a spine with both localized activity (top) and activity that invades the entire branch (bottom). **a_2**, automatic detection of ROIs on the same dataset using d-NMF identifies the same spine with local (top) and global (bottom) activity. **b_1**, Also from Sheffield & Dombeck, *Nature* 2015, illustration of a stretch of dendrite of a mouse CA1 pyramidal cell expressing GCaMP6f showing large area branch spiking and a putative locally-generated spike. **b_2**, automatic detection of ROIs using d-NMF identifies similar ROIs as those manually identified and accurately detects global and local activity. **c_1**, From Cornejo et al., *Science* 2022, a dendritic branch of a pyramidal neuron in mouse somatosensory cortex expressing the voltage indicator postASAP, showing localized activity in numbered spines and adjacent dendrites. **c_2**, extracted ROIs (left) and corresponding activity (right) from d-NMF on the same dataset. **d_1**, Maximum projection image of a mouse CA3 pyramidal neuron from this study. The red box indicates the zoomed-in view of a region containing a dendritic spine (**d_2**) automatically identified by d-NMF as having localized activity. **d_3**, The activity of this spine at times is synchronous with the nearby stretch of dendrite (global, top) but also shows localized activity (local, bottom).



Reviewer Fig. 1 | Fine to coarse segmentation of dendrites based on merging threshold. Each row represents one dataset from Fig. 7. The raw output of d-NMF consists of many ROIs with slightly different activities. By setting progressively lower merging thresholds, the user can obtain the level of detail required for their scientific question.

Q3. *It seems entirely plausible that SNR and event duration may vary as a function of the dendritic compartment (basal vs apical dendrites) and dendritic distance from the soma and that these may explain some of the CA3 results presented here.*

I still find it difficult to understand how downsampling to a rate of 1.5 Hz does not hinder the ability to detect fast (~200 ms long) localized dendritic spikes (as well as the performance of the algorithm when combined with other, faster indicators). The authors have demonstrated that there were similar ROIs detected in the full session (not-downsampled) and the downsampled session (Extended Data Fig. 2); however, they have not demonstrated the constraints of SNR, event frequency, and event duration on their ability to detect localized transients. One approach to demonstrate this is to leverage their techniques for synthetic data (used in Fig. 3): artificially insert events into localized dendritic regions with a range of spatial extents, signal amplitudes, frequencies and durations and demonstrate the bounds by which the algorithm performs.

Similarly, extended Data Fig. 3 demonstrates the ability to separate spatially contiguous dendrites that have varying levels of independence, but the SNR for dendritic compartments appears to be unreasonably high, and the degrees of independence for which the algorithm can successfully identify independent ROIs are not shown. Consider the (very realistic) case where the full dendritic branch is activated synchronously throughout a recording session, except for one or two localized (~ 10 um) events of small amplitude (as has been described in hippocampal neuron dendrites in vivo). Will d-NMF be able to identify such events and divide this branch into smaller ROIs? With the synthetic data in extended Fig. 3, exploring these bounds should be straightforward.

A3-1. The reviewer raises an important point, and once again provides excellent suggestions that have further prompted us to add more useful features to our algorithm. The central concern is that the act of downsampling may systematically exclude small, rapid events in the dendrites and that this effect may bias results. Directly to this point, we have repeated all of our scientific analyses on non-downsampled 30 Hz data to test the effect of inclusion of thin transients. This extensive effort revealed two important findings:

1) First, using 30 Hz data we observe that apical and basal dendrites have a higher activity rate than soma, which differs from our original claim that basal dendrites had lower rates than the soma **(Fig. 6f**). This is likely due to the inclusion of rapid dendritic calcium transients that we were not detecting in the downsampled data. We are very grateful to the reviewer for this instructive suggestion.

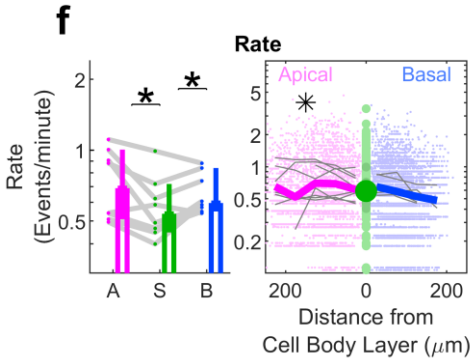


Fig. 6 | f, Left, apical and basal dendrites had a significantly higher mean transient rate than the soma, but were not different from each other (Apical: 0.69, [0.51, 1.01] Events/minute; Soma: 0.54, [0.45, 0.72] Events/minute; Basal: 0.60, [0.55, 0.84] Events/minute; Apical vs Soma, p=0.016; Apical vs Basal, p=0.38; Soma vs Basal, p=0.039, Wilcoxon sign-rank test, n=8 sessions). Right, apical, but not basal, dendrites showed increased transient rates as a function on distance from the cell body layer (Apical: $p=1.1\times10^{-8}$, two-way ANOVA; Basal: p=0.20, two-way ANOVA).

2) Second, the results of spatially tuned apical dendrites being more stable across days (**Fig. 7**) and functioning as better position decoders (**Fig. 8**) remain statistically significant using 30 Hz data, further validating the findings.

A3-2. At the reviewer's request, we also plotted the SNR and event duration (transient width) as a function of distance from the soma. Both apical and basal dendrite show thinner transient widths compared to the soma; in the apical dendrites transient widths also decreased as a function of distance from the soma (**Fig. 6g, Supplementary Fig. 13e**). There are small statistically significant differences in SNR between dendrites and soma, but these do not affect our scientific conclusions (**Supplementary Fig. 11d, j**).

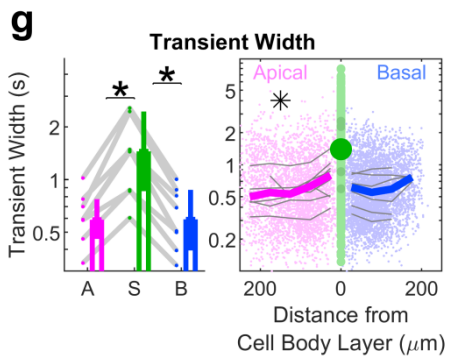
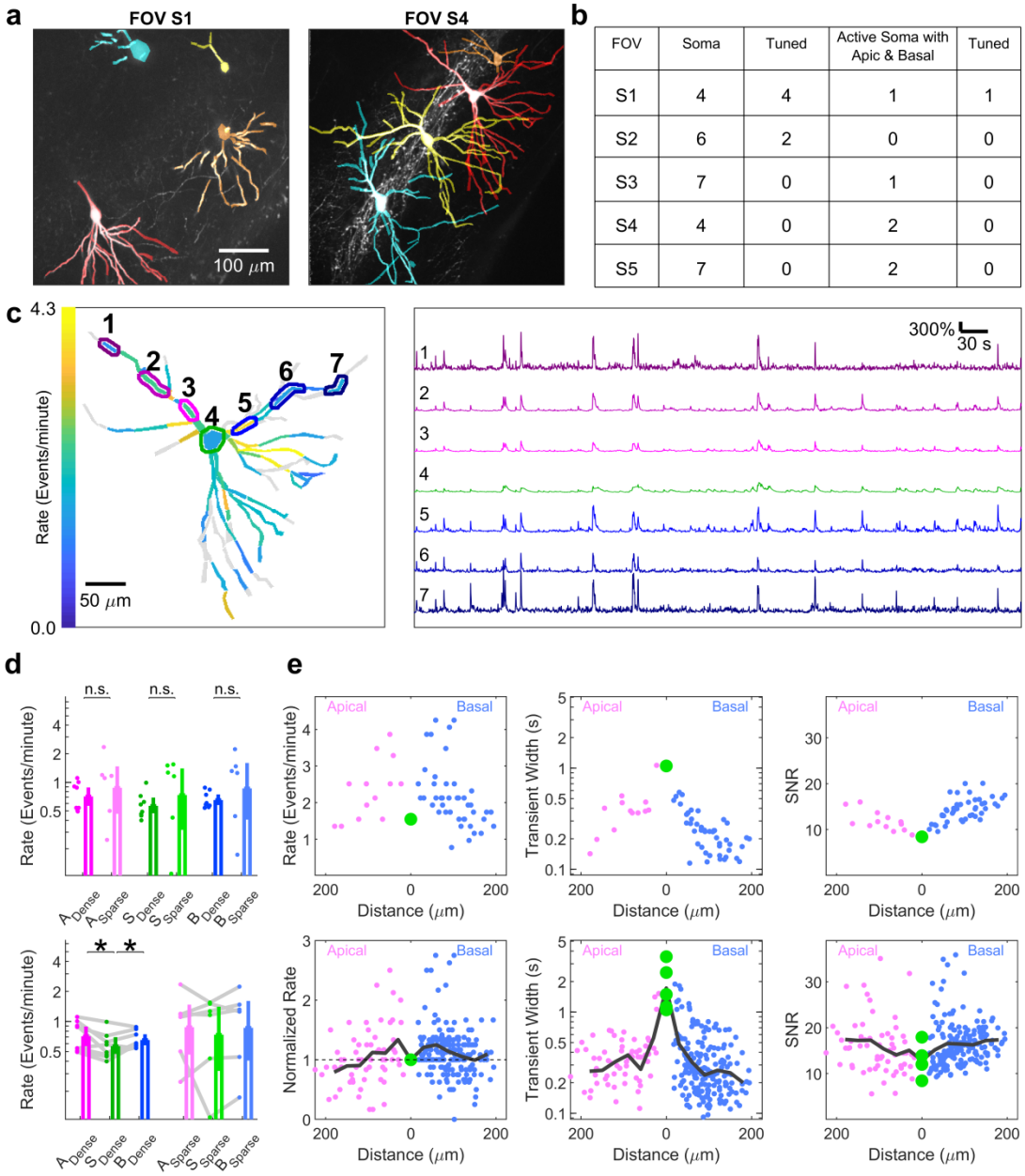
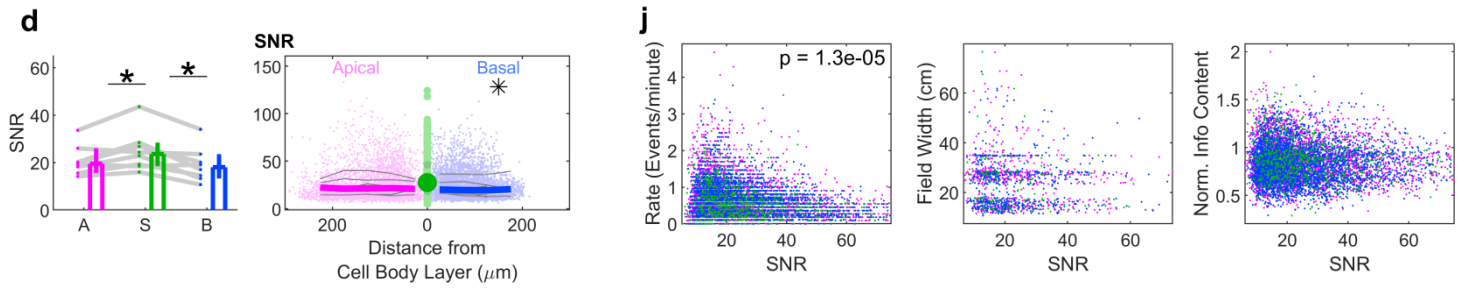


Fig. 6 | g, Left, the half-width of calcium transients was smaller in apical (590, [460, 770] ms) and basal (590, [390, 870] ms) dendrites compared to soma (1400, [860, 2400] ms; Apical vs Soma, p=0.0078; Apical vs Basal, p=0.95; Soma vs Basal, p=0.0078, n=8 sessions, Wilcoxon sign-rank test for all). Right, apical dendrites showed a decrease in transient width with increasing distance from the cell body layer (apical, $p=3.3\times10^{-9}$, two-way ANOVA) but not basal dendrites (basal, p=0.48, two-way ANOVA).



FOV	Soma	Tuned	Active Soma with Apic & Basal	Tuned
S1	4	4	1	1
S2	6	2	0	0
S3	7	0	1	0
S4	4	0	2	0
S5	7	0	2	0

Supplementary Fig. 13 | Sparsely labeled data reproduces dense dataset trends. a. Maximum intensity projections of two sample fields of view from sparsely labeled mice, with neuron skeletons overlaid in different colors. **b.** Table showing the number of soma obtained in each field of view. **c.** Left, sample neuron (red highlighted neuron from FOV S4) pseudocolored with the event rate in each compartment. 7 sample ROIs are indicated with outlines. Right, fluorescence traces for the 7 indicated ROIs. Note there are more calcium transients the closer the ROI is to the soma (ROI 4, green). Activity is plotted as ΔS/S, where S is the ratio of green GCaMP to red tdTomato (see Methods). **d.** Top, overall activity rates were comparable between dense and sparse data for apical dendrites (Dense: 0.71, [0.57, 0.88]; Sparse: 0.88, [0.47, 1.48], p=0.35, Wilcoxon rank-sum test), soma (Dense: 0.57, [0.47, 0.70]; Sparse: 0.74, [0.32, 1.40], p=0.28, Wilcoxon rank-sum test), and basal dendrites (Dense: 0.65, [0.58, 0.74]; Sparse: 0.86, [0.40, 1.60], p=0.35, Wilcoxon rank-sum test). Bottom, apical and basal dendrites had significantly higher rates than soma in the dense data. This trend was observed in the sparse data as well (Apical vs Soma, p=1; Apical vs Basal, p=0.56; Soma vs Basal, p=0.22, Wilcoxon sign-rank test for all). Data here are presented as mean and 95% confidence interval of the mean rather than median because there are only 6 data points in the sparse dataset, making the median a brittle measure. **e.** Top, activity rate, transient width, and signal-to-noise ratio (SNR) as a function of distance from the soma for the example neuron in **c.** Bottom, aggregated normalized rates, transient widths, and SNR for all recorded neurons in the sparse datasets. Black line indicates the average in 30 μm bins.



Supplementary Fig. 11 | Additional quantification of basic place tuning properties. d, Left, signal-to-noise ratio (SNR) of apical (20, [16, 26]) and basal dendrites (18, [13, 24]) were smaller than soma (24, [18, 28]; Apical vs Soma, p=0.039; Apical vs Basal, p=0.078; Soma vs Basal, p=0.016, n=8 sessions, Wilcoxon sign-rank test for all). Right, basal dendrites showed a decrease in transient width with increasing distance from the cell body layer (basal, p=0.049, two-way ANOVA) but not apical dendrites (apical, p=0.48, two-way ANOVA). **j,** Controlling for SNR, there was a significant effect of compartment type on event rate ($p=1.3x10^{-5}$, 3-way ANOVA, see Methods), but not field width (p=0.054) or normalized information content (p=0.47).

A3-3. Downsampling has its advantages in terms of improving SNR and accelerating the time for data processing/analysis without imposing high demands on computing capacity and sacrificing the number of ROIs detected. However, it may not be the optimal choice for all, depending on the given scientific question (eg. ensemble coding versus detection of fast and small transients characteristic of local events). Temporal downsampling is not a requirement but a choice we leave to the user, and in the newest version of the algorithm, we additionally allow the user to adjust the amount of temporal and spatial smoothing performed on the video. These are useful "knobs" to adjust based on the data at hand and the specific scientific question posed by the user. This is illustrated in the revised Fig. 2a.

From the main text, lines 222-231: Higher sampling rates are generally required to detect rapid events. However, a higher sampling rate comes at the cost of noisier data, which may adversely impact ROI detection and refinement. Indeed, in our hippocampal data, performing the *Refinement* step on 30 Hz data resulted in slightly lower coverage, though the overall performance against manually labeled data was relatively similar (Supplementary Fig. 9) to temporally downsampled data. Additionally, processing raw data increases both the size of files in memory and computation time. Depending on the expected kinetics and spatial extents of activity as well as the computing power available, users may choose to downsample or filter in time, space, or both. Because we are investigating dendritic activity, which may be rapid, all following analyses are done on data extracted from non-downsampled, 30 Hz files.

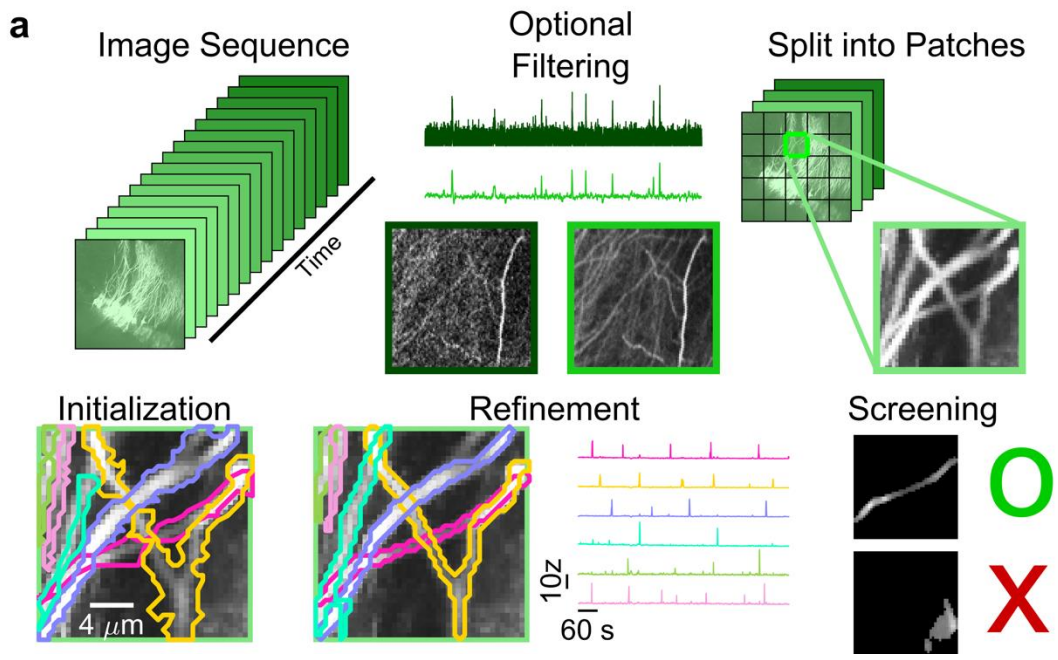
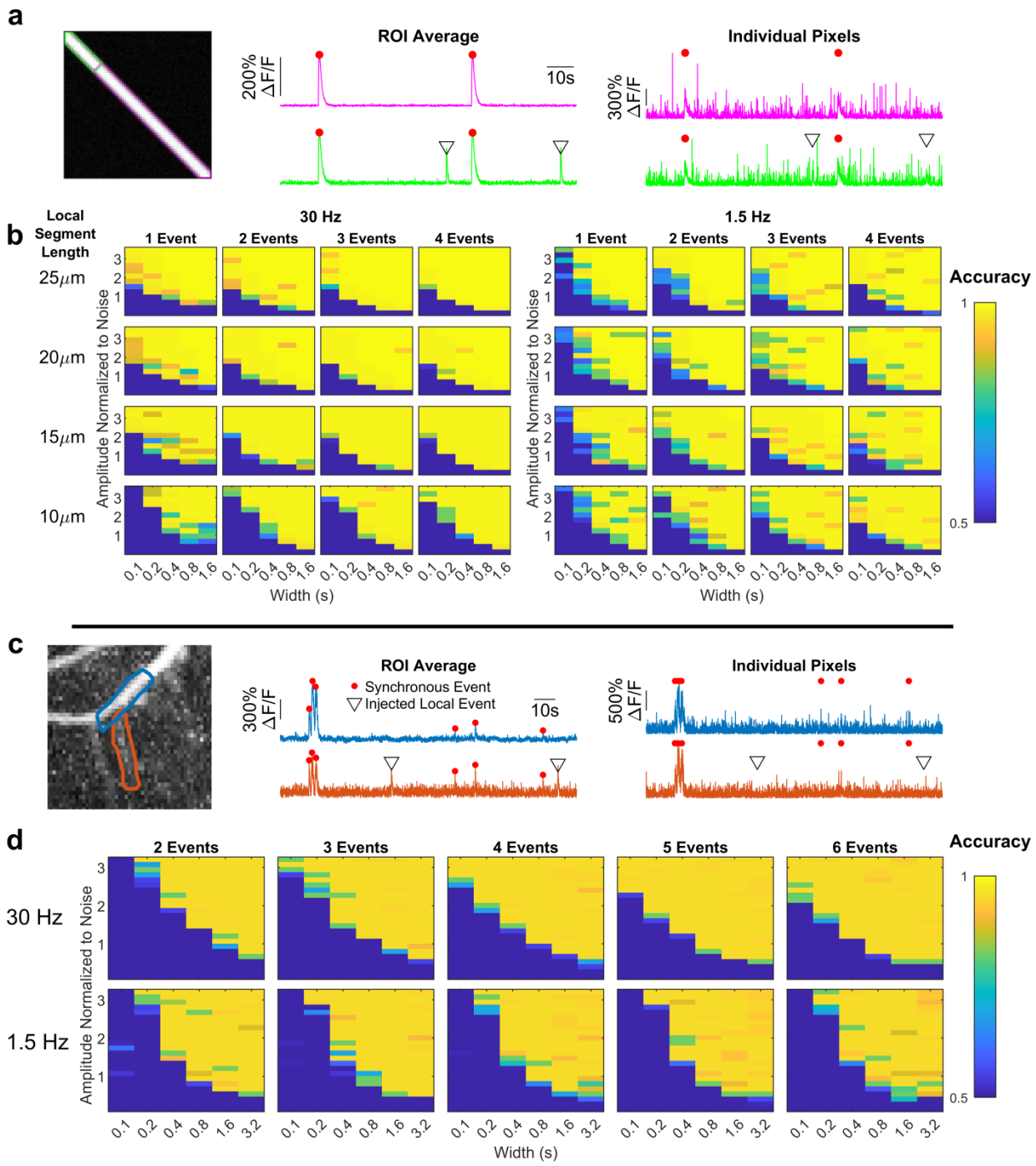


Fig. 2 | d-NMF pipeline. a, Pictorial representation of the workflow. Image sequences may be filtered and/or downsampled to reduce memory requirements and improve signal-to-noise ratio. The downsampled image sequence is then split into patches. Within each patch, ROI cores are detected and their spatial footprint and temporal traces are iteratively estimated using sparse constrained NMF. Once all patches have been processed, overlapping ROIs from neighboring patches are tested to see if they can be merged. Finally, ROIs are screened for realistic activity (see panel **c**).

A3-4. Finally, based on the reviewer's suggestions, we constructed two artificial scenarios to test the limits of the algorithm in detecting rapid and localized transients. We prepared synthetic datasets by injecting calcium transient waveforms of different amplitude, duration, and number into localized dendritic regions of interest with known baseline activity. This analysis demonstrated that in non-downsampled, 30 Hz data, localized transients as brief as 100 ms in width were sufficient to identify independent ROIs, provided the transients were of sufficient amplitude to rise above the noise level (**Supplementary Fig. 8**). We additionally verified that we were detecting such thin transients in our real data (**Fig. 5d**).

From the main text, lines 214-220: To explore the bounds of activity in which d-NMF could detect localized activity, we prepared synthetic datasets by injecting calcium transient waveforms of different amplitude, duration, and number into localized dendritic regions of interest with known baseline activity. This analysis demonstrated that by analyzing non-downsampled, 30 Hz data, localized transients as brief as 100 ms in width were sufficient to identify independent ROIs, provided the transients were of sufficient amplitude to rise above the noise level (Supplementary Fig. 8).



Supplementary Fig. 8 | Exploration of limits of d-NMF in detecting localized dendritic activity. a, Left, schematic of synthetic dataset arrangement. Two ROIs (magenta and green) are defined on a single dendritic branch. Synchronous events are present in both ROIs, while local events are present only in the green ROI. The parameters varied were the length of the green ROI, the amplitude of local events, the width of local events, and the number of local events. Each parameter combination was simulated 3 times. Middle, average fluorescence activity of the two ROIs for a given simulation (2 local events, local length of 20 pixels, amplitude of 1.6 x noise level, width of 800 ms). Right, fluorescence activity for a randomly-selected single pixel of each ROI. Note that while the mean trace (middle) shows a high signal-to-noise ratio, individual pixels are much noisier. **b**, Left, accuracy of d-NMF when run at 30 Hz in detecting the localized activity as a function of width and amplitude, and number of the local events as well as the length of the independent segment. Accuracy is defined as the mean number of ROIs detected across all simulations, divided by 2. Provided local events are large enough (>1.5 x noise levels), even very brief localized events can be detected. Right, same as left but when the signal was first downsampled to a rate of 1.5 Hz. **c**, Schematic of a second synthetic dataset arrangement, in which localized calcium transients were injected into a dendritic branch (red outline) of a real dataset. As in **a-b**, the

number, duration, and amplitude of localized events were systematically varied, and the ability of d-NMF to detect these localized transients were tested. Each combination of parameters was simulated twice. d, As in b, the accuracy of d-NMF in detecting localized activity. Accuracy here is defined as the highest spatial correlation of all detected ROIs with the branch ROI outlined in red in **c**. If the branch ROI was not treated as independent, then this value will be low. The maximum value across 5 simulations is displayed. The top row shows the accuracy when run on 30 Hz data. The bottom row shows the accuracy when the data was downsampled to 1.5 Hz.

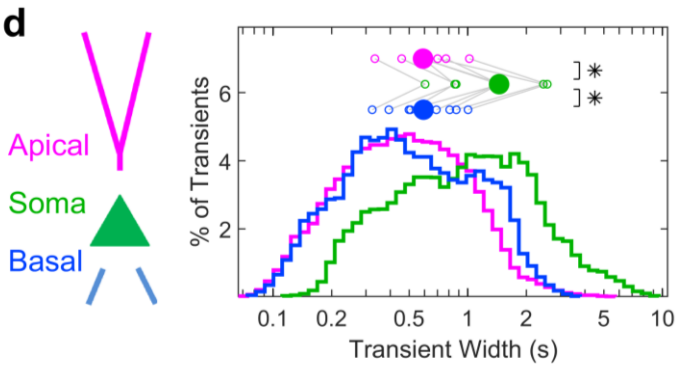


Fig. 5 | Fitness trace screens out false activity. d, Left, Schematic designating apical dendrites, soma, and basal dendrites. The half-width of detected calcium transients spanned 2 orders of magnitude in apical dendrites, basal dendrites, and soma, with a bias towards wider transients in the soma. Lines represent histograms of events averaged over sessions. Open circles indicate the mean value of all transients within individual sessions. Filled circles indicate the median value across sessions. Apical: 590, [460, 770] ms; Soma: 1400, [860, 2400] ms; Basal: 590, [400, 870] ms. Apical vs Soma: $p=7.8\times10^{-3}$; Apical vs Basal: $p=0.95$; Soma vs Basal: $p=7.8\times10^{-3}$. N=8 sessions, Wilcoxon sign-rank test for all.

Reviewer #2

Q4. *The authors claim that my comments were "orthogonal" to the goals of the manuscript and thus did not address them. This manuscript presents a lot of work and that is impressive in itself. The second half of the manuscript could even be presented separately as a scientific study (and maybe it should). However, I don't think that diminishes the concerns I have specifically about the computational method, and the need to critically evaluate this part *as* a computational method.*

To repeat, the two main problems I see with such dendrite segmentation methods are the ***backpropagating AP*** *and the* ***oversplitting of ROIs****. The authors do not address the backpropagating AP, and claim that oversplitting does not change statistical tests. The choice not to address my comments has directed me to look more carefully at the design of the study and I have identified two new issues.*

A4-1. Firstly, as we did in the previous response statement, we would like to acknowledge that the points raised by the reviewer are conceptually and scientifically important considerations in the dendritic imaging field. We certainly did not choose not to address these concerns. We tried to demonstrate that merging ROIs with similar activity does not change scientific conclusions compared to the ROIs originally detected by our algorithm as a good-faith effort to address and explore this issue. We sincerely apologize to the reviewer that our response did not satisfactorily address the concerns raised. In the revised text and below points, we make an effort to more thoroughly address them.

(1) Detecting **local dendritic events** and dissociating them from **bAPs**

(a) We demonstrate in 4 independent data sets (**Supplementary Fig. 7**) that our algorithm can detect localized transients, even thin and small amplitude ones typical of local dendritic events

(b) We verified that our algorithm works with a voltage imaging data set, a game-changing tool under development for all-optical resolution of sub-threshold, local and bAP-driven dendritic activity *in vivo*

(c) We assert that while our algorithm is capable of detecting localized dendritic events and resolving them from bAPs, our current data set and experimental design are not optimized for studying bAPs. We discuss why this remains an important challenge to solve in the field, but does not preclude one from making exciting discoveries about dendritic function

(2) Concerns regarding **over splitting**

(a) We establish that DNMF does not systematically oversplit and provides comparable-sized dendritic ROIs to existing analysis software (Suite2P), and several published studies using various approaches to ROI segmentation.

(b) We elaborate on our method of algorithm evaluation and demonstrate that our performance metric does not depend on the number of ROIs detected.

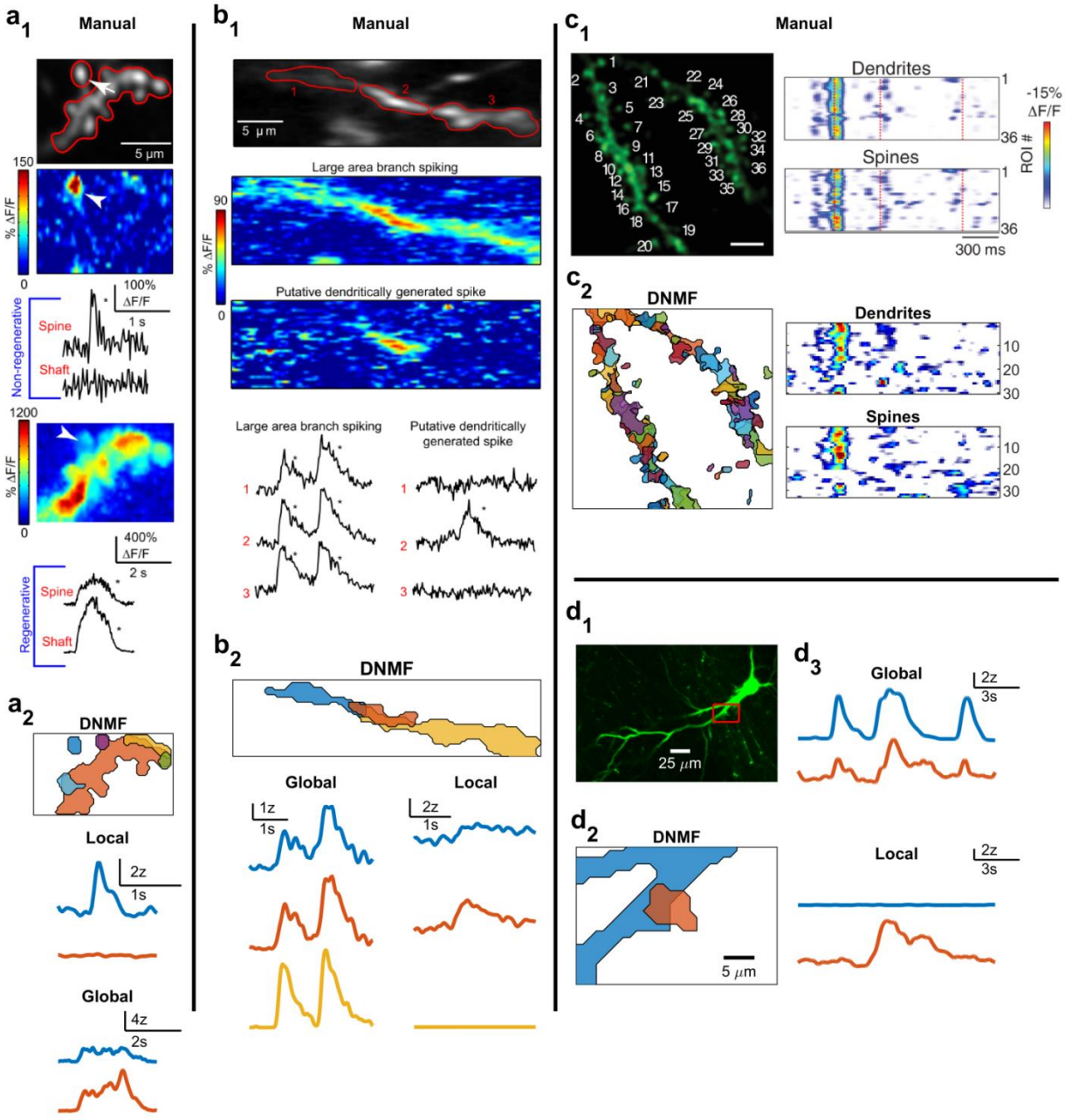
(c) We provide users with a variable merging threshold for ROI segmentation based on experimental purpose and scientific query;.

A4-2. Resolving local dendritic events from backpropagating APs *in vivo* is indeed an important research direction. The reviewers raised a key related question about whether our algorithm can detect localized and relatively fast Ca^{2+} transients in dendrites characteristic of local events (Reviewer 1 and Reviewer 2), and the answer is yes. Not only can our algorithm detect local dendritic Ca^{2+} signals, but it can also handle voltage imaging data, thereby opening up new avenues for future studies to interrogate soma-dendrite coupling dynamics and dissociating the directional origin of dendritic activity (forward vs. back-propagating).

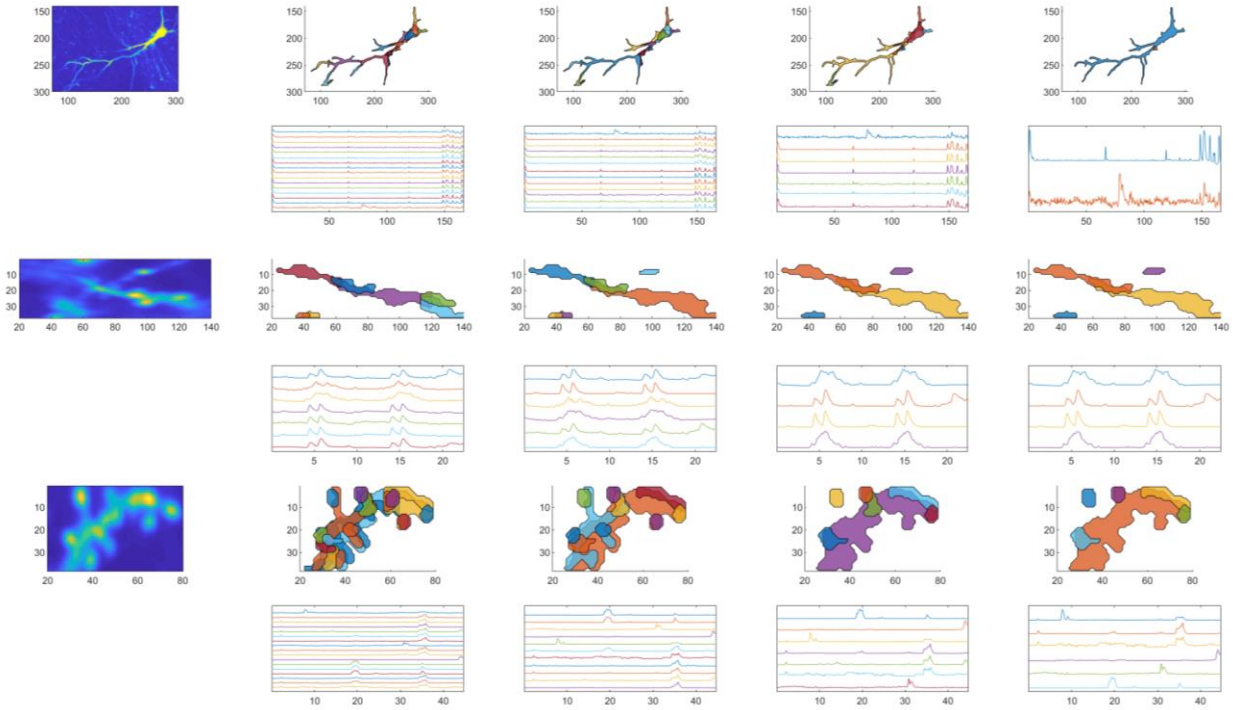
See also our response to **Q2**, replicated below:

> ***"Q2.*** *However, I am still concerned about the ability of d-NMF to detect localized dendritic transients. Clarifying this point is critical for future users, as well as validating the experimental results presented in the manuscript. Low SNR and fast timescale localized dendritic events have previously been suggested to play a role in the hippocampus in memory formation.*
>
> **A2.** We agree that the ability to detect localized dendritic transients in a wide variety of preparations is important for our approach to enjoy widespread use and success. We have taken this opportunity to demonstrate our method on 4 separate new datasets from 2 other pioneering labs that are experts in the field of dendritic two-photon imaging as well as experimental and synthetic data from our own lab. Each data set shows instances of localized dendritic activity that our method detected and extracted.
>
> From the main text, lines 210-214: We validated that d-NMF was able to detect localized activity from two previously published datasets, using calcium imaging in mouse hippocampal area CA1[21] and voltage imaging in visual cortex[38], respectively. We also identified local activity in a putative dendritic spine in our own data (Supplementary Fig. 7).
>
> To arrive at these ROIs, we adjusted the threshold at which to merge ROIs. This threshold can be adjusted by the used depending on how fine- or coarse-grained the user desires. This is illustrated in **Reviewer Fig. 1**."

Editorial Note: Content in panels a and b reproduced from https://doi.org/10.1038/nature13871, with permission from Springer Nature. Content in panel c reproduced from https://doi.org/10.1126/science.abg0501 with permission from AAAS.



Supplementary Fig. 7 | Fully automatic Identification of localized dendritic activity by d-NMF. a_1, From Sheffield & Dombeck, *Nature* 2015, illustration of a stretch of dendrite of a mouse CA1 pyramidal cell expressing GCaMP6f showing a spine with both localized activity (top) and activity that invades the entire branch (bottom). **a_2**, automatic detection of ROIs on the same dataset using d-NMF identifies the same spine with local (top) and global (bottom) activity. **b_1**, Also from Sheffield & Dombeck, *Nature* 2015, illustration of a stretch of dendrite of a mouse CA1 pyramidal cell expressing GCaMP6f showing large area branch spiking and a putative locally-generated spike. **b_2**, automatic detection of ROIs using d-NMF identifies similar ROIs as those manually identified and accurately detects global and local activity. **c_1**, From Cornejo et al., *Science* 2022, a dendritic branch of a pyramidal neuron in mouse somatosensory cortex expressing the voltage indicator postASAP, showing localized activity in numbered spines and adjacent dendrites. **c_2**, extracted ROIs (left) and corresponding activity (right) from d-NMF on the same dataset. **d_1**, Maximum projection image of a mouse CA3 pyramidal neuron from this study. The red box indicates the zoomed-in view of a region containing a dendritic spine (**d_2**) automatically identified by d-NMF as having localized activity. **d_3**, The activity of this spine at times is synchronous with the nearby stretch of dendrite (global, top) but also shows localized activity (local, bottom).



Reviewer Fig. 1 | Fine to coarse segmentation of dendrites based on merging threshold. Each row represents one dataset from Fig. 7. The raw output of d-NMF consists of many ROIs with slightly different activities. By setting progressively lower merging thresholds, the user can obtain the level of detail required for their scientific question.

The scientific conclusions of our study about place dendrites and the increased stability and decoding accuracy of apical dendrites are robust and impactful, without relying on resolving bAPs, and now stand validated in two independent data sets (sparse and dense) and sampling rates (downsampled vs. full rate). **Reviewer Table 1** below provides examples of imaging studies of dendritic activity published across various journals, generating valuable insights about dendritic function without explicitly delineating local dendritic events from backpropagation, compared to the few that had to combine *in vivo* imaging with electrophysiology to examine local dendritic vs. bAP signals.

Beyond analyses, directly resolving bAPs and local dendritic activity requires deliberately designed experimental approaches, such as simultaneous somatic and dendritic electrophysiology with well-defined stimulation protocols, and tool development such as *in vivo* optimized voltage imaging sensors. The technical challenges involved have limited progress in this direction, and only two published studies have been able to dissociate dendritically generated local events from bAPs *in vivo* (Jia et al. 2010; Cornejo et al., 2022). For us to address this question sufficiently would represent a sizeable experimental expansion of the manuscript, which, as the reviewer acknowledges, is already fairly extensive. Hence, this is experimentally out of the scope of the paper. We intend to pursue this exciting question in future collaborative studies.

in vivo dendritic imaging and electrophysiology studies dissociating bAPs

Konnerth Lab - A study in the mouse primary visual cortex, recorded sub-threshold and suprathreshold activity from pyramidal neuron soma *in vivo* with patch clamp electrophysiology, while simultaneously imaging somatic and dendritic responses using a Ca2+ indicator loaded intracellularly through the patch pipette (Jia et al., 2010). This study correlated activity in dendritic hotspots (3 µm dendritic ROIs) with bAPs from the soma due to the combined electrophysiology and imaging approach at the single neuron level. However, this is extremely technically challenging and not feasible in the context of studying sparsely represented tuning features of dendrites and soma.

Furthermore, this study highlights the need to split the dendritic compartments to a fine scale in order to resolve local events independent of bAPs. Studies in mouse auditory and barrel cortex, showing adjacent spines of pyramidal neurons may be tuned to different tone frequencies and whisker stimulations, respectively (Chen et al., 2011; Varga et al., 2011), also make a case in favor of fine spatial resolution for ROIs.

Yuste Lab- Cornejo et al., Science 2022- Two-photon voltage imaging of dendrites with simultaneous soma recordings in the primary visual cortex L2/3 pyramidal neurons reveals backpropagation action potentials (bAP), a rich repertoire of dendritic spikes, and spine activation without somatic depolarization. This suggests that the activity of individual spines was not detectable in the soma.

Dendritic compartment specific activity restricted to 3-5 µm.

in vivo dendritic imaging studies without resolving bAPs

Gan Lab- Cichon and Gan, Nature 2015: different branches of the same motor cortical cell are tuned to either forward or backward motion.

The average dendritic ROI length in this study is 10 µm.

Dombeck Lab- Sheffield and Dombeck, Nature 2015 and Neuron 2017: branch prevalence of Ca2+ signal predict the precision and formation of CA1 place fields, respectively.

In the 2015 study, dendritic shaft segments where putative localized dendritically generated Ca2+ events were resolved averaged ~10µm in ROI length, while global large area branch spiking spanned ~30 µm across the dendrite length. Non-regenerative 'local' Ca2+ events were restricted to spines of 2 µm ROI span.

Harnett lab- Beaulieu-Laroche et al Neuron 2019: semi-simultaneous dendritic and somatic imaging in the awake primary visual cortex suggests strong functional correlation between somatic and dendritic compartments, regardless of neural activity levels, visual stimuli, or locomotion

Svoboda Lab – Kerlin eLife 2019: Most dendritic calcium transients coincided with global events, while task-associated calcium signals in dendrites and spines were compartmentalized by dendritic branching and *clustered within branches over approximately 10 µm.*

Smirnakis lab- Park et al., Nature Comm. 2019: 2-photon laser micro-dissection on layer-2/3 pyramidal neuron dendrites in mouse primary visual cortex highlighted the resilience of orientation tuning to dendritic input loss. Removing the apical dendritic tuft nor basal dendrites did not alter orientation-tuning curves, and ablation of at least 2 basal dendrites was needed to cause a small shift in orientation preference.

Rochefort Lab - Francioni et al eLife, 2019: Imaged GCaMP6s signals in tuft dendrites, trunk and soma of layer 5 pyramidal neurons in mouse primary visual cortex. The signals in the apical tuft were highly coupled to trunk and somatic transients, but the frequency of calcium transients decreased a s a function of distance from the soma to the tuft

Losonczy Lab – O'Hare et al., Science 2022: Basal dendrites are more co-tuned with the soma compared to apical dendrites, and elevating intracellular Ca2+ release by genetically deleting an ER-mitochondrial tethering protein increases the level of spatial co-tuning of apical dendrites with their soma in CA1 place cells, and widen and stabilize place fields.

Please note to analyze intradendritic Ca2+ activity dynamics and spatial feature selectivity within a dendritic branch, each dendritic ROI was split into 2µm sub-ROI segments to detect isolated Ca2+ transients.

Schiller Lab – Otor et al., Science 2022: Tuft dendrites of early bifurcating /large nexus pyramidal tract neurons (PTNs) show branch-independent activity representing different motor variables, whereas late bifurcating /small nexus PTNs dendrites show homogenous activity.

The ROI lengths for dendrites in this study range from 15-50 µm, which is comparable to the range of ROIs our algorithm generates.

Luthi Lab- De'Aquin et al. Science 2022: In pyramidal neurons in Lateral Amygdala during spontaneous and an auditory cued fear conditioning behavior, learning induced plasticity is decoupled between soma and dendrite due to modulation by compartment specific inhibition

Reviewer Table 1 | Overview of previous studies investigating dendritic activity using optical methods

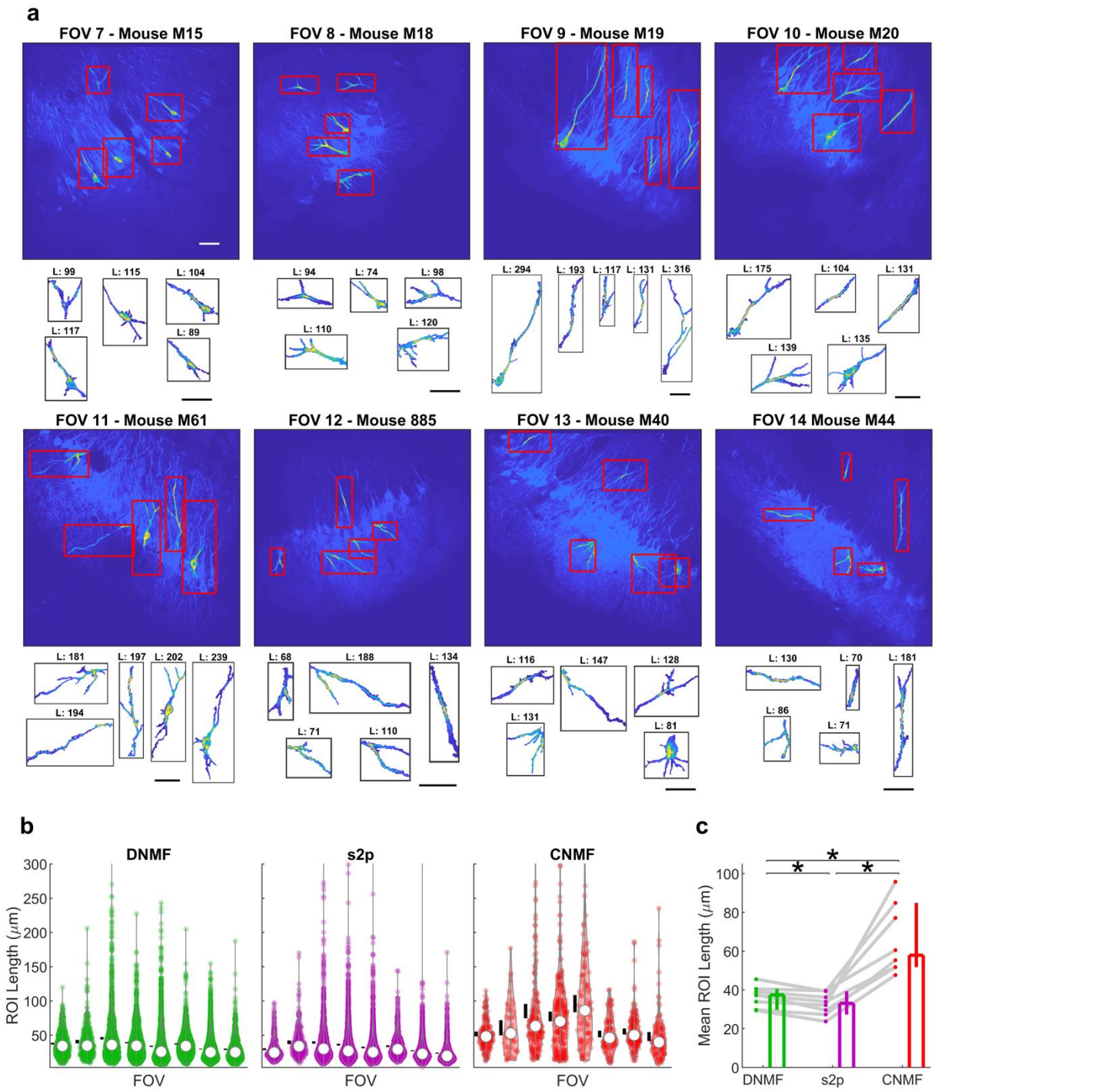
Q5-1. *Not only is d-CNMF oversplitting ROIs, but the human annotators making the ground truth have themselves been directly instructed to oversplit ROIs, which is described in the methods:*

"Additionally, labelers were instructed to break up long dendritic branches into smaller ROIs about the length of the soma (approximately 30 pixels)."

I don't think this is ok to do. Breaking the dendrites into pieces of approximately this length will make the ground truth more like some algorithms (presumably more like d-CNMF) but this will not constitute a meaningful ground truth comparison. IF the ground truth is biased in this way, then my second new issue arises: older methods were not optimized to extract "soma-length" segments of dendrites. In fact, I believe both CNMF and Suite2p are trying to find segments as long as possible. These older algorithms have also not been optimized by the authors for the number of ROIs returned, even though some of their main parameters control this (number of clusters in CNMF and threshold in Suite2p).

A5-1a. This is an important point, and we appreciate the opportunity to address this concern by explaining our evaluation metrics in more detail. Note that the following points are relevant specifically to the evaluation of the algorithms; a discussion of the statistical implications is included further below. First, we will note that d-NMF does not systematically oversplit ROIs. The algorithm was not tuned to find ROIs of a certain size, only ROIs that are contiguous in space and have homogenous activity. The below figure (**Supplementary Fig. 5**) illustrates several examples from our dense fields of view in which identified ROIs span great distances and incorporate multiple branches. Furthermore, the length of ROIs detected by d-NMF is slightly but significantly longer than the length of ROIs identified by suite2p in the same datasets.

From the main text, lines 201-204: The success of d-NMF compared to other methods is not due to systematic detection of a higher number of smaller ROIs; ROI size was largely comparable across methods, with several instances of ROIs spanning well over 100 µm identified (Supplementary Fig. 5).

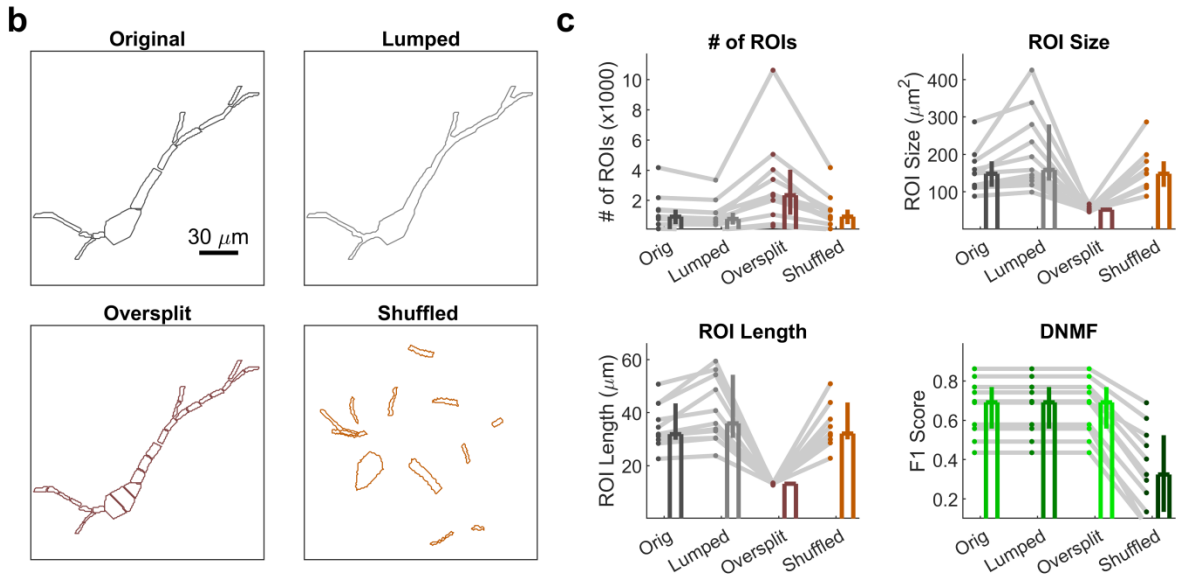


Supplementary Fig. 5 | Illustration of large ROIs detected by d-NMF. a, 8 fields of view from our dense datasets, with 5 sample ROIs each highlighted. Each ROI is labeled with its length L, defined as the longest path (in μm) between any two member pixels. 1 pixel corresponds to 1.085 μm; scale bars denote 50 μm. b, Distributions of ROI lengths for the 8 dense datasets from DNMF, suite2p, and CNMF. **c**, The median ROI length for DNMF was 37, [30, 41] μm, significantly larger ($p=7.8x10^{-3}$, Wilcoxon sign-rank test) than ROIs obtained from suite 2p (s2p) (33, [27, 39] μm) and significantly smaller ($p=7.8x10^{-3}$, Wilcoxon sign-rank test) than those obtained from CNMF (58, [52, 85] μm). CNMF ROIs were also significantly larger ($p=7.8x10^{-3}$, Wilcoxon sign-rank test) than suite2p ROIs.

A5-1b. Second, we note that **our performance metrics are not based on the number of ROIs identified**. The metric of the number of ROIs is perhaps appropriate for fields of view composed of soma, in which one ROI is equivalent to one soma. But as the reviewer rightly points out, one could game the system by splitting a dendrite (or soma for that matter) into a large number of arbitrarily small ROIs. Furthermore, we wanted to use a metric insensitive to how split up an ROI was. Put another way, if one set of ROIs used 30 pixel segments on average and another used 50 pixel segments, we wanted to treat each clustering as equally valid, provided the entire neuron was labeled by some ROI.

The performance metrics we report in Fig. 4, then, are purely pixel-based, not ROI-based. A "True Positive" represents a pixel that belongs to any ROI in the ground truth set and test set. A "False Positive" represents a pixel assigned to an ROI in the test set but not the ground truth set. A "False Negative" represents a pixel assigned to an ROI in the ground truth set but not the test set. These values are summed together and combined to calculate the F1 score and True Positive Rate reported in **Fig. 4**.

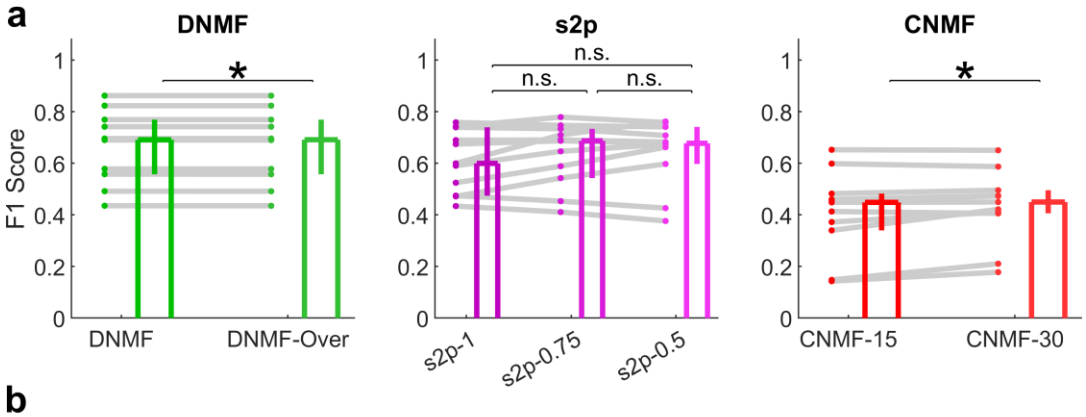
As a result, these evaluation metrics should not depend on the size (and consequently number) of ROIs in the ground truth data, but rather on their proper placement. We verified this in three ways, summarized in **Supplementary Fig. 4b-c**. First, we manually went through all ground truth data and merged ROIs that plausibly belonged to the same neuron or dendritic branch, giving a "Lumped" ground truth dataset. Second, we took those lumped ROIs and split them up into 10 um long segments, giving an "Oversplit" dataset. Finally, we took the original ROIs and shuffled them by rotating and translating each ROI independently, yielding the exact same size and number of ROIs but in the wrong positions. As expected, the evaluation metrics for d-NMF, suite2p, and CNMF were identical between the Original, Lumped, and Oversplit ground truth datasets, while the methods compared to the Shuffled ground truth dataset had lower evaluation scores.



Supplementary Fig. 4 | Evaluating ROI segmentation performance. b, Demonstration that the precise segmentation of an ROI is irrelevant provided that correct pixels are covered. The "original" data is made of ROIs as they were originally labeled. "Lumped" data merges ROIs that are spatially overlapping and have temporally correlated. "Oversplit" data splits the original ROIs into segments approximately 10 um in length. "Shuffled" data is obtained by shifting and rotating the original ROIs independently by random amounts. **c,** Though the number of ROIs, their size, and length change across manipulations, the overall F1 score is unchanged except when compared to the shuffled dataset. **Number of ROIs:** Original, 930, [430, 1400] ROIs; Lumped, 780, [400, 1200] ROIs; Oversplit, 2300, [1100, 4000] ROIs; Shuffled, 930, [430, 1400] ROIs. **ROI Size:** Original, 150, [110, 180] μm^2; Lumped, 160, [130, 280] μm^2; Oversplit, 52, [48, 55] μm^2; Shuffled, 150, [110, 180] μm^2. **ROI Length:** Original, 32, [30, 43] µm; Lumped, 36, [30, 54] µm; Oversplit, 13.2, [12.9, 13.4] µm; Shuffled, 32, [30, 44] µm. **F1 Score:** Original, 0.69, [0.56, 0.77]; Lumped, 0.69, [0.56, 0.77]; Oversplit, 0.69, [0.56, 0.77]; Shuffled, 0.32, [0.13, 0.52].

From the main text, lines 184-190: Because ROIs may be broken up into a different number of segments, labeling completeness was evaluated on a pixel-by-pixel basis, with true positives indicating that a pixel belongs to at least one ROI in both the manually-labeled and automatically labeled ROI sets (see Methods). Note that measuring performance in this manner quantifies the degree of proper coverage of the dataset, and is insensitive to the total number of ROIs detected or how they may be split up (Supplementary Fig. 4b).

A5-1c. The reviewer is correct that we did not try to optimize the other algorithms to match (or maximize) the number of ROIs returned. In response, we re-ran suite2p with 2 different values for the "threshold" parameter (0.75 and 0.5 compared to 1 in the original version), and re-ran CNMF with a higher value of "number of clusters" (30 compared to 15 in the original version). We also performed an artificial oversplitting of the output of d-NMF by splitting ROIs into sub-ROIs of length 10 µm, as described above. As shown in the below figure (Reviewer Fig. 2), oversplitting did not affect the evaluation of d-NMF at all. Lower values of "threshold" did result in a modest increase in suite2p performance. CNMF was not substantially affected by the modification. It should be noted that the output of CNMF is screened by a classifier of "valid" and "invalid" ROIs, so the number of ROIs returned is not exactly twice that of the previous version.



b

Measure	DNMF	DNMF-Over	s2p-1	s2p-0.75	s2p-0.5	CNMF-15	CNMF-30
F1	0.69	0.69	0.60	0.69	0.68	0.45	0.45
# ROIs	1190	3012	812	1507	2165	109	164
ROI Size	142.8	43.9	99.8	84.8	79.4	436.4	414.9
ROI Length	34.3	12.0	29.6	28.0	23.5	49.2	46.7

Reviewer Fig. 2 | Adjusting thresholds for segmentation methods does not substantially affect accuracy. a, Left, the F1 score for DNMF was practically unchanged if ROIs were oversplit into 10 µm-long subsegments (DNMF, 0.69, [0.56, 0.77]; DNMF-Over, 0.69, [0.56, 0.77]; p=0.02, Wilcoxon sign-rank test). Statistical significance here is due to small decreases in F1 score due to small ROIs being discarded. Middle, suite2p had a small gain in F1 score when the "threshold" parameter was lowered from 1 (0.60, [0.47, 0.74]) to 0.75 (0.69, [0.54, 0.73]) to 0.5 (0.68, 0.60, 0.74]) s2p-1 vs s2p-0.75, p=0.15; s2p-1 vs s2p-0.5, p=0.46; s2p-0.75 vs s2p-0.5, p=0.76, Wilcoxon sign-rank test). Right, CNMF output was not substantially more accurate when the "number of clusters" parameter was changed (CNMF-15, 0.45, [0.34, 0.48]; CNMF-30, 0.45, [0.41, 0.50]; p=0.03, Wilcoxon sign-rank test).

In summary, we have demonstrated that 1) Our method does not systematically produce ROIs that are smaller than other methods; 2) Our evaluation criteria depend on the correct placement of ROIs, and not the total number of ROIs; and 3) Altering the number of ROIs in other methods does not substantially affect their accuracy. These are important points and we have made them more prominent in our description of the algorithm and our evaluation criteria.

Q5-2. *Even though the authors did not directly present how much oversplitting from d-CNMF there is in the real data, we can infer it from the new analyses using simulations: at the 1.5 quality threshold with 50-100 neurons in the field of view, you are oversplitting neurons by a factor of 20-40 (!!) while at the more conservative 3.0 quality threshold it is still a factor of 5-10.*

A5-2a. We believe there may have been a misunderstanding here and are sorry if we did not clearly communicate the methods and results of the new simulation figure (**Fig. 3**). The figures below should clarify. First, we took a sparsely labeled field of view containing a neuron with a long dendritic tree with many branches and manually labeled all processes. This includes dendrites of the primary neuron along with nearby axons. This resulted in 105 manually drawn ROIs. Of those, 78 belonged to the primary neuron and 27 belonged to other processes. We constructed simulated data by copying the movie, rotating and shifting it in space and shifting it in time. Every copy then adds in an additional 105 ROIs. In **Fig. 3e** we quantify how many manually drawn ROIs have a signal quality above a certain threshold. If all ROIs had sufficient signal quality then the # of ROIs would be 105*(# Neurons). The fact that the observed curves fall well below this line indicates that signal quality is being degraded by the addition of uncorrelated activity from the shifted movies. However, to avoid confusion, we have remade the figure and quantified the total pixel area belonging to ROIs that are of sufficient quality, rather than the total number of ROIs, with the message remaining similar: as more neurons are added to a FOV, signal quality degrades and some ROIs are unrecoverable from noise.

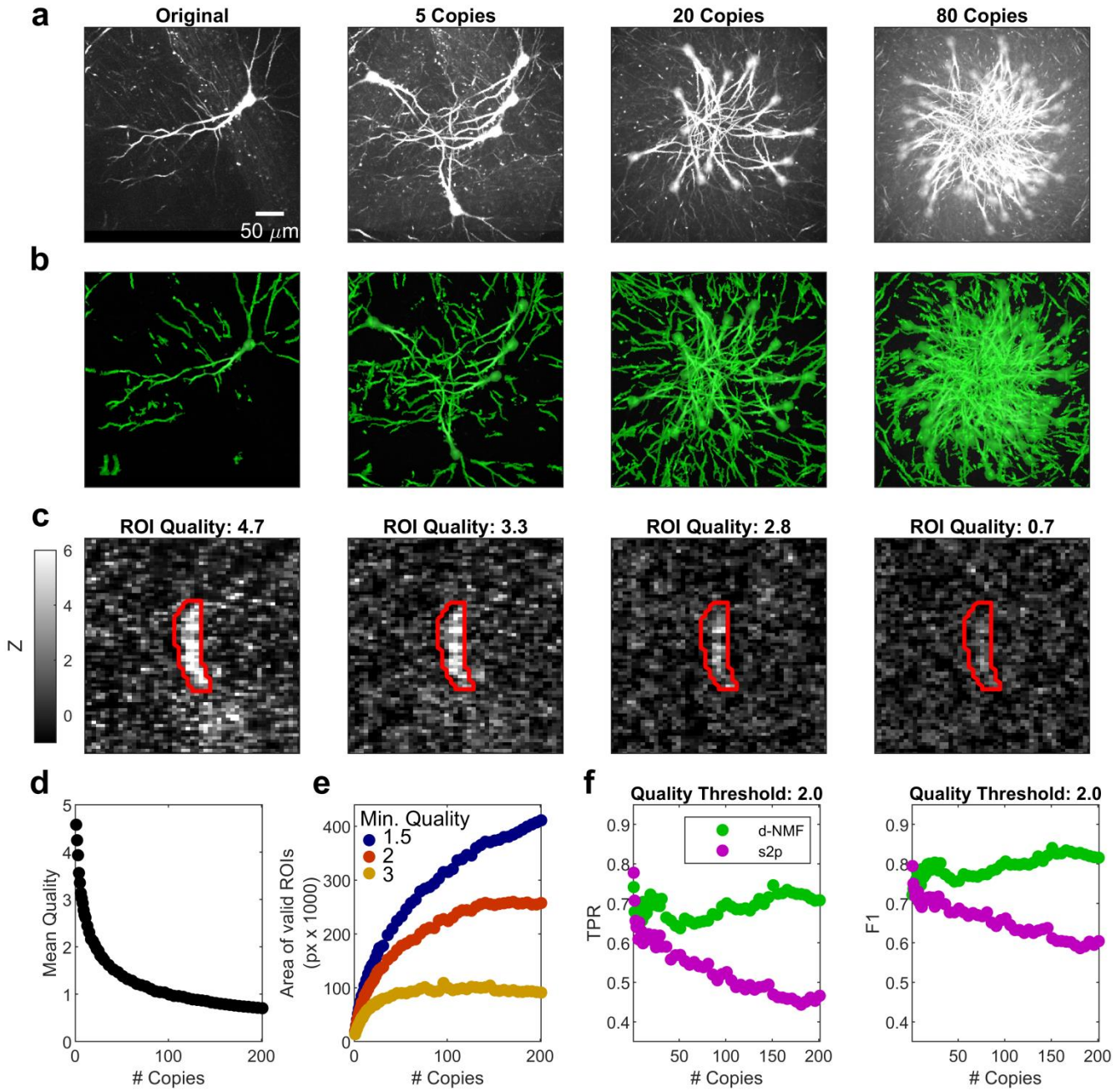
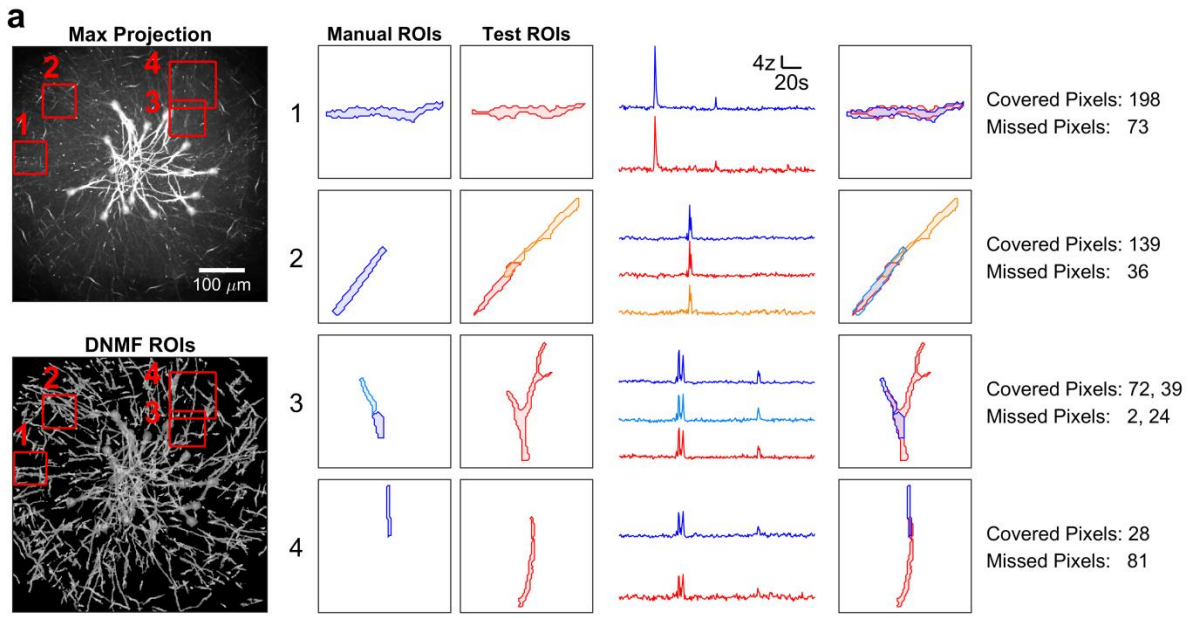


Fig. 3 | Synthetic data demonstrates that d-NMF remains accurate at high density. a, Maximum projection images of movies generated by copying, rotating, and shifting the movie in the first column. **b**, ROIs derived from d-NMF for the corresponding movie in **a**. **c,** Frame of maximum activation for a sample ROI in the corresponding movies in **a**. The signal quality of this ROI is noted in the titles, and decreases as more neurons are added to the synthetic data. **d**, The mean signal quality of ROIs decreased as more neuron replicates were added to the synthetic data. Quality is defined as the mean z-scored value of all pixels at the ROI's most active frame. **e**, The area of ROIs with signal quality above a certain threshold (here 1.5z, 2z, and 3z) initially increased linearly but slowed down or decreased at higher densities. **f**, Left, restricted to ROIs with a signal quality >2z, true positive rate for d-NMF (green) and suite2p (s2p, purple) as a function of the number of neurons in the synthetic dataset. Right, F1 score for the same data. The d-NMF method maintained high performance even for very dense datasets.

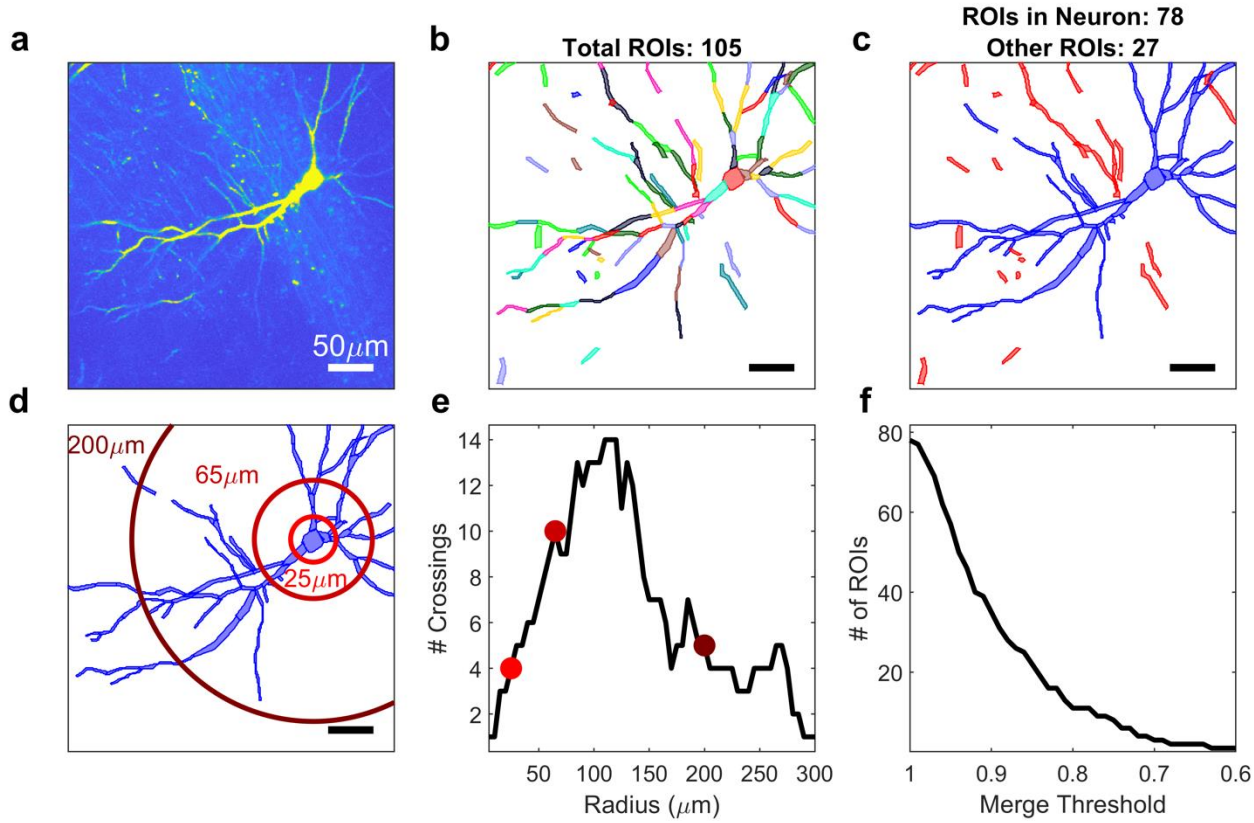
A5-2b. The robustness of d-NMF as an algorithm is demonstrated in **Fig. 3f**. Unlike in the dense data, we not only know the ground truth ROIs, but we also know the ground truth activity of each ROI. This allows us to evaluate the algorithms

in a slightly different way. Rather than evaluating correctness on a pixel-by-pixel basis, we now match ROIs that have similar activity traces between the ground truth set and the algorithm's set. This is described in the methods and summarized pictorially in the **Supplementary Fig. 4a**. This analysis shows that not only are we correctly identifying ROI boundaries, but we are correctly estimating their activity. This evaluation is only possible in this synthetic dataset because we know the ground truth activity.



Supplementary Fig. 4 | Evaluating ROI segmentation performance. a, Sample field of view from the synthetic dataset (top) and ROIs identified from d-NMF (bottom). Insets show provide examples of how ROI segmentation performance is evaluated. Manually labeled ROIs are indicated in blue shades, and d-NMF identified ROIs ("test ROIs") are indicated in red shades. A manual ROI is scored by the proportion of pixels that are covered by any test ROI that has similar activity.

A5-2c. Even if we restrict discussion to ROIs belonging only to the primary neuron, a relevant question is if 78 ROIs is too many. Performing a modified Sholl analysis (restricted to 2 dimensions rather than 3 because we image only a single plane) shows that this neuron's dendritic tree is highly branched, peaking at 14 branches ~100 um from the cell body (**Reviewer Fig. 3**). We would assert that assigning the entirety of this neuron to a single ROI would result in a drastic loss of information. To be able to investigate branch-specific activity, each branch must be assigned to its own ROI. The question then comes down to how many branches and whether or not there is branch-specific activity, which can be evaluated by clustering ROIs based on the correlation of their activity. We show that as the threshold for merging decreases the number of ROIs also decreases. For the entire neuron to be considered a single ROI we would have to use a very permissive merging threshold of 0.6. Small amplitude, fast time-course independent events are expected to have only a marginal effect on correlation values, so we prefer to use a higher correlation cutoff to enable the possibility of analyzing local activity. But this threshold for merging is left to the user.



Reviewer Fig. 3 | Segmenting the sparse neuron template. a, Maximum-intensity projection of the neuron duplicated for Fig. 3. b, Manual segmentation of all 105 ROIs in the field of view. c, ROIs color-coded according to those that are connected to the neuron of interest (78 ROIs, blue) and those that are not (27 ROIs, red). d, Illustration of the modified Sholl analysis method, showing radii of 25, 65, and 200 µm. e, The number of crossings as a function of radial distance from the cell soma, peaking at 14 crossings. Radii marked in panel d are marked with corresponding circles. f, The number of ROIs obtained based on the correlation threshold used to merge similar ROIs. A permissive threshold value of 0.6 would be needed to merge all ROIs within the neuron into a single ROI, which would eliminate the ability to investigate localized differences in activity.

Q5-3. *It cannot be mathematically true that oversplitting does not affect statistiscal analyses. If an analysis has multiple copies of the same replicates (5-20 copies of each replicate in this case), then the statistiscal tests will by definition be incorrect, since the assumptions of independence are strongly invalidated. No merging analysis can change this basic statistical fact.*

A5-3. We would like to clarify that the statistical tests use the number of Fields of View (8 FOVs) as the N, rather than individual ROIs. That is, for each FOV, we compute each measure for ROIs individually and then take their mean within ROI type (apical, soma, basal) to arrive at a single number for each ROI type for each session. Note that if each ROI was exactly replicated 5 times this would not affect the mean estimates for the sessions, as duplicate values would not introduce any bias, and the statistical test between compartments would return the same because the number of FOVs did not change.

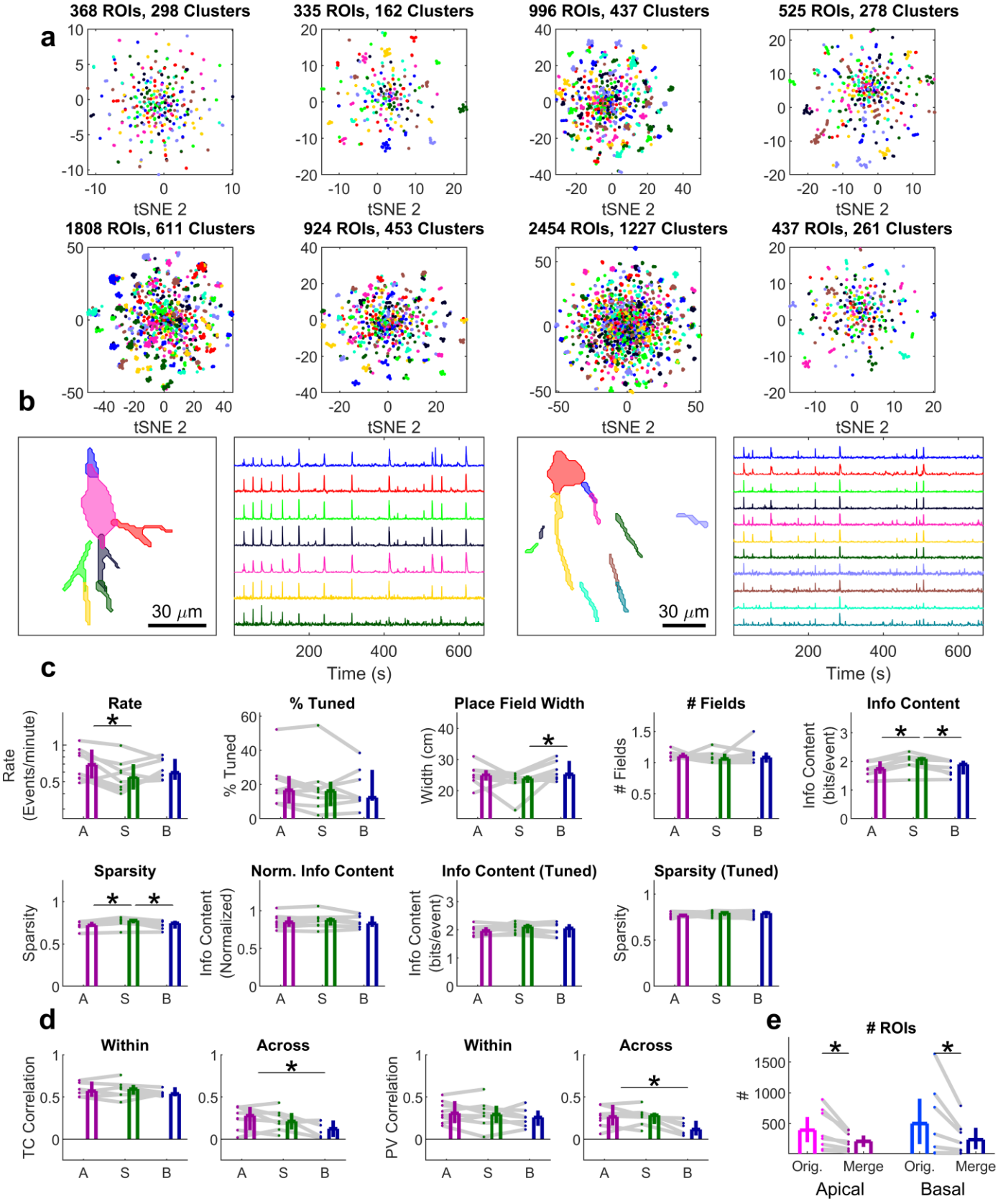
This confusion may have arisen from our wording. The title of **Supplementary Fig. 12** (Previously Supplementary Fig. 11) was previously "Oversplitting does not affect statistical measures," and we stated "All statistical comparisons were identical in the original and merged datasets" in our previous response letter. What we meant to say was that the scientific conclusions are not affected. A view of our table of statistics for that figure (Supplementary Table 4, compared with Fig. 6 and Supplementary Fig. 11) makes it clear that the precise statistical values do differ between the two conditions. The updated title for this figure is "Merging ROIs with similar activity does not affect scientific conclusions"

Q5-4. *Further, the amount of merging for this analysis seems minor (only 30% reduction in number of apical dendrites, whereas their simulations show that at least 80% reduction is necessary, and sometimes 97% reduction depending on data quality and SNR). The authors could claim that the simulations are designed to push the limits of the method, but 50-100 neurons in a field of view seems relatively normal and the oversplitting seems even worse at lower neuron numbers. Still, the amount of oversplitting in real datasets can be easily gaged by running more sophisticated clustering methods, and visualized with dimensionality reduction methods, like tSNE and UMAP.*

A5-4. Thank you for the great suggestion for visualizing ROI relationships using such methods. We took the reviewer's advice and did a tSNE visualization in our dense datasets, and indeed can see clusters of ROIs that have similar activity that likely belong to a single neuron. The plot below uses a hierarchical clustering on the correlations between ROI activity with a cutoff value of 0.7. But as we stated in a previous point, we feel this decision is best left up to the user to set the threshold at which to merge similar ROIs. Small differences in activity between two ROIs may be noise that one wishes to average over, or they may reflect bona fide localized, independent activity. If we choose to merge two ROIs their activity is by definition synchronous and such independence is lost.

We have repeated the merging analysis on our non-downsampled, 30 Hz data, and the results are very comparable between merged and non-merged data. Dendrites have higher activity rates than soma, though the difference between basal dendrites and soma does not reach statistical significance in the merged dataset. Importantly, the stability results of apical dendrites being more stable than basal dendrites remains in the merged dataset, illustrating the robustness of this finding.

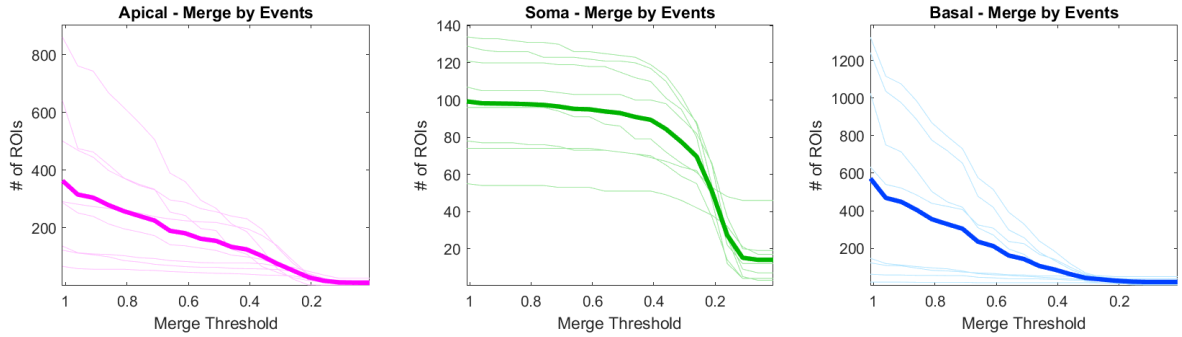
From the main text, lines 308-311: Additionally, these results were not biased by "oversplitting" of dendrites into multiple compartments, as all comparisons were similar when we merged ROIs with highly correlated activity (see Methods, Supplementary Fig. 12).



Supplementary Fig. 12 | Merging ROIs with similar activity does not affect scientific conclusions. a, Cluster plots for each FOV in the dense dataset, visualized by t-distributed stochastic neighbor embedding (tSNE). Each dot represents a single ROI, and ROIs are clustered according to their temporal correlation with each other (see Methods). **b,** Two sample cluster groups showing neurons that have been split up into multiple ROIs. **c,** Statistical measures show similar trends and significant differences between the original data and data obtained from merging

ROIs with similar activity. See Table 4 for summary statistics and compare with Fig. 6 and Supplementary Fig. 11 for values measure on the original, non-merged data. **d,** As in **c,** but for stability measures. Compare with Fig. 7 for values measured on the original, non-merged data. **e,** The number of apical dendrites in the merged dataset (210, [120, 310], n=8 FOVs) was ~45% smaller than the number in the original dataset (390, [190, 610]). Similarly, the number of basal dendrites in the merged dataset (240, [80, 430]) was ~50% smaller than the number in the original dataset (500, [160, 900]).

Based on the above explanation (**A5-2**) we feel like an 80-97% reduction in the number of ROIs is not a useful target. For our dense datasets, we computed the number of ROIs obtained for different threshold levels for merging. To match the number of soma on average, we would have to merge ROIs at a correlation value of ~0.35, which is very small indeed. The goal is not to arrive at a single ROI each for the apical dendrite, soma, and basal dendrite for each neuron, but to assign ROIs to portions of the dendritic branch that have sufficiently correlated activity.

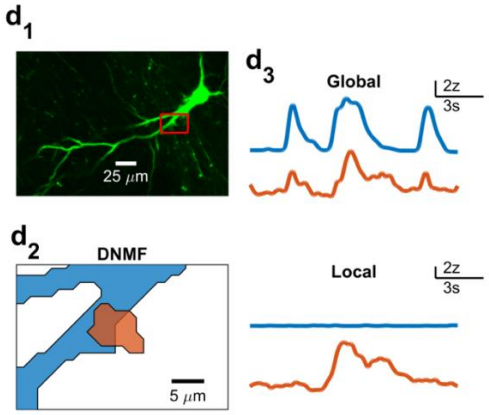


Reviewer Fig. 4 | Number of ROIs as a function of merge threshold. The number of ROIs across all FOVs is plotted as a function of merge threshold for apical dendrites, soma, and basal dendrites. Thin lines represent individual FOVs, and thick lines represent the median.

Q6. *For investigating the effect of the BAP, you can evaluate whether any activity exists in local dendritic compartments that is significantly different from the rest of the dendritic tree. Do any of your ROIs look like spines? If there isn't any localized dendritic activity, then we're looking at different versions of the BAP in different parts of the dendritic tree. Which is fine and potentially interesting, but it does not really address questions about dendritic integration and computation.*

A6. See responses **A4-2** and **A2** above. To summarize, we have included new analyses from our data and that of others showing that activity in local dendritic compartments can indeed be detected by our algorithm, particularly in putative spine compartments (**Supplementary Fig. 7d**). Importantly, such an analysis would not be possible if the entire neuron or branch was assigned to a single ROI. It can only be done if the dendritic tree is split into several compartments; otherwise there would just be a single global signal that we would not be able to localize to a particular spine or branch.

A thorough and systematic analysis identifying precisely which calcium events are local versus back-propagated and the relative contribution of each to coding properties would indeed be fascinating, but it would require substantial experimental and conceptual additions to the manuscript. While we are highly interested in such questions, we are inclined to agree with the reviewer that this is best left to a separate manuscript.



Supplementary Fig. 7. d_1, Maximum projection image of a mouse CA3 pyramidal neuron from this study. The red box indicates the zoomed-in view of a region containing a dendritic spine (**d_2**) automatically identified by d-NMF as having localized activity. **d_3**, The activity of this spine at times is synchronous with the nearby stretch of dendrite (global, top) but also shows localized activity (local, bottom).

Q7. *If my concerns are not seen as relevant here, then I am not sure what the method is innovating on. Is this method really an improvement over existing approaches like Suite2p and CNMF, which also give you a lot of oversplit dendritic ROIs without considering the effect of the BAP, and which allow you to control how many ROIs you want to get? The new method is claimed to recover a higher fraction of ROIs, but the older methods were not using optimized parameters for these particular datasets. Simple changes such as increasing the number of components (CNMF) and reducing the threshold (Suite2p) would have increased the number of returned ROIs and thus the matching to the "ground-truth" ROIs, but these analyses were not done by the authors.*

In summary, the authors have not addressed my original two concerns (oversplitting and BAP) and their comments led me to identify two new concerns (artificially over-segmented ground truth and failure to optimize previous algorithms to this artificial scenario).

A7. We do consider the reviewer's concerns as relevant and, as stated above, in the previous round we had generated a figure (**Supplementary Fig. 12**, Previously Supplementary Fig. 11) dedicated to addressing oversplitting. In addition, we discussed reasons why a thorough examination of BAP was beyond the scope of this study. We are sorry if our explanations were not clear enough to address the concerns satisfactorily. The concern of statistical estimates being biased due to oversplitting is very important. Here, we present multiple lines of analysis (**Supplementary Figs. 4, 5, 12)** to demonstrate that neither our algorithmic evaluations nor our scientific conclusions are a direct outcome of oversplitting. These additional checks should be considered standard in future dendritic studies.

Regarding the number of ROIs obtained by different methods, we were careful to use evaluation methods that did not depend on the number of ROIs obtained, and we have made that more obvious in the revised manuscript. Increasing the number of ROIs did give a slight performance boost to suite2p, but not for CNMF. While suite2p and d-NMF are comparable in terms of identifying ROI outlines, d-NMF provides better estimates of the true activity of these ROIs.

We would like to reiterate that the effects of BAP are indeed very important and deserving of careful study. But the first steps to understanding those effects are to detect dendritic ROIs and arrive at accurate estimates of their activity, which is what we achieve here.

In summary**,** we have demonstrated that neither over-segmented ground truth datasets nor an increase in the number of ROIs returned by previous algorithms meaningfully impact the results due to the evaluation metrics we use. We have expanded our initial investigation into the impacts of oversplitting on our scientific results and have demonstrated that we are able to detect localized activity in dendritic spines from genetically encoded Ca^{2+} as well as voltage signals, which lay the tools and groundwork to deliberately study bAPs in the future. Systematically separating local activity from BAP and evaluating the contribution of each on spatial coding properties remains an important goal, but we prefer to use the space allotted in this manuscript to demonstrating the validity of the computational method.

Reviewer #3:

Q8. *the authors have addressed all of my concerns.*

A8. We thank the reviewer for their helpful comments which have made the manuscript better.

We have formatted this response as follows:

Original remarks from the reviewers in black italics

Our responses in blue

Quotations from the updated manuscript in red

Remarks and responses are numbered by Query (Q1, Q2, Q3…) and Answer (A1, A2, A3…)

Reviewer #1 (Remarks to the Author):

Q1. The authors have nicely addressed our previous comments. The manuscript presents an algorithm that has been optimized for defining ROIs over dendrites; there are novel initialization and refinement steps and I feel that many researchers in the community will be users. The authors also present new findings related to the functional properties of CA3 dendrites that I find interesting and compelling.

A1. We sincerely thank the reviewer for their comments, implementing which has made the study stronger.

Q2. I have also been asked to comment on the most recent review from another reviewer:

Related to bAPs, oversplitting and how to merge dendritic segments: The authors argue that merging of ROIs is up to the user, which I largely agree with. It is not clear what the best method is to merge ROIs, and this process could be quite different for users with different types and quality of data. In fact, the underlying biology related to in vivo dendritic function is largely unknown, which further implies that it is premature to develop an algorithm that claims to have the correct merging procedure. This does not mean that an algorithm without such a capability is not useful. For example, with a particular merging algorithm implemented, it would be possible to look for differences in dendrites between conditions. Whether the algorithm is the "correct" one or not would, in such a case, not prevent authors from detecting interesting dendritic functional changes across conditions.

However, the authors do some merging analysis in their rebuttal using an off-the-shelf hierarchical algorithm, and I agree with the reviewer that this should be included in the manuscript to give readers an idea of how such a step could be accomplished. Such a section in the manuscript could be expanded and explored more with the synthetic data analysis that that authors have used so nicely up to this point. As presented, the authors use a merging threshold and demonstrate that their scientific results hold for a range of thresholding values. Such robustness analysis is valuable to the techniques presented, but I believe that the reviewer's concern could be largely addressed by emphasizing in the text that such robustness analysis is necessary for users to perform on their own.

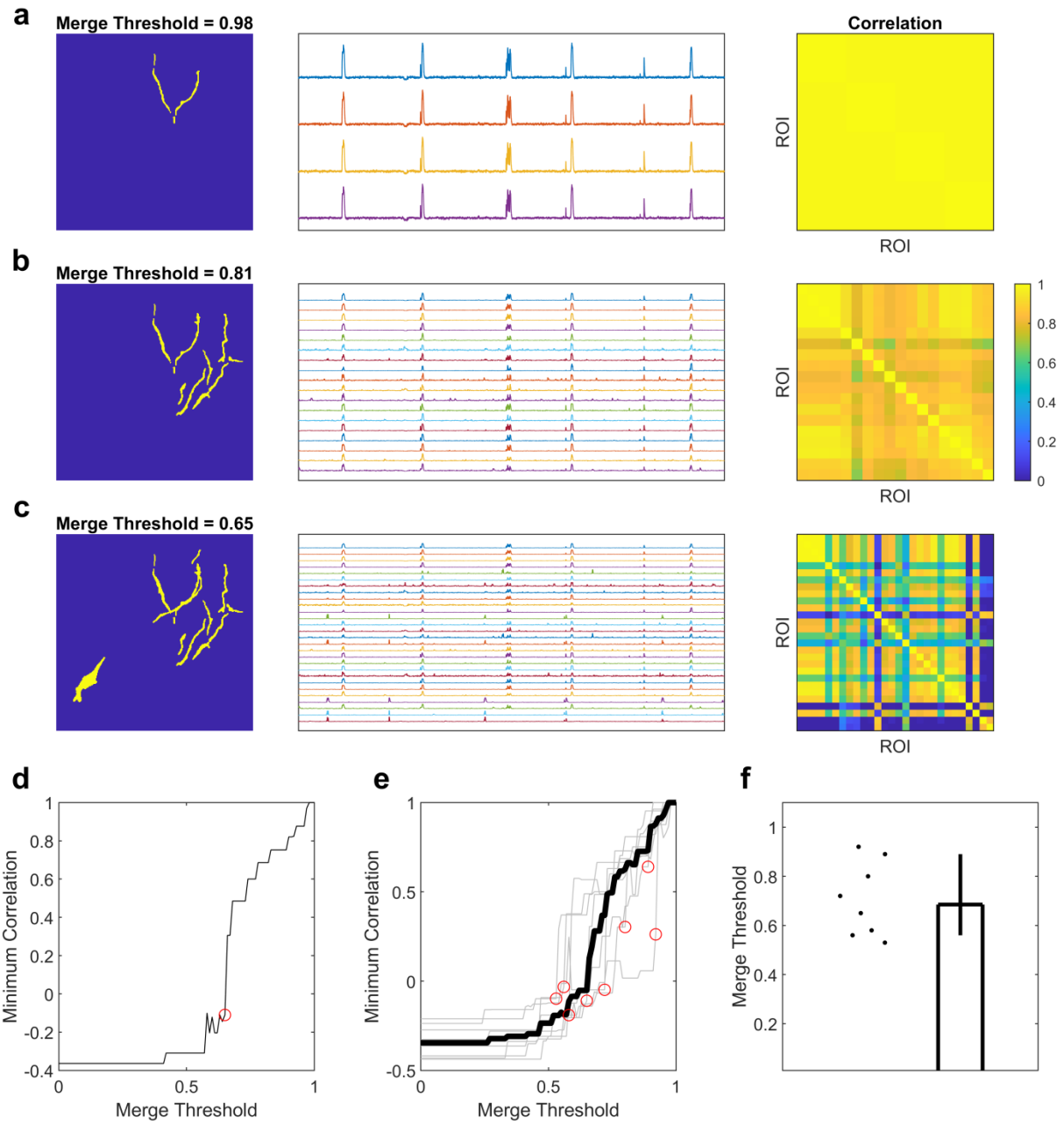
A2. We are happy that the reviewer agrees with us on this point. The question of defining a set of criteria with which to merge ROIs is important, and likely varies across datasets. As per both reviewers' suggestions, we have now included the figures exploring the effects of different merging thresholds in the manuscript **(Supplementary Figs. 3, 10)**, and we include sections in the **Results (Lines 153-162)**, **Discussion (Lines 415-427)**, and **Methods (Lines 673-685)**, emphasizing that users must take an active approach in curating the output of our algorithm.

Results (Page 4, Paragraph2, Lines 153-162):
"In the Merging step, we combine ROIs (Supplementary Table 2c) that are likely to correspond to the same functional unit. Long dendritic branches may span multiple image patches (Fig. 2a) or may be represented by multiple ROIs due to local variations in activity. The algorithm offers users the ability to set a threshold correlation with which to merge ROIs. A high threshold can result in a fine parcellation of dendritic structure, preserving local differences in activity, while a low threshold will merge ROIs with broadly the same activity. We observed that a general rule on the lower bound of the merging threshold is to identify when ROIs not belonging to the same neuron become linked together. This is

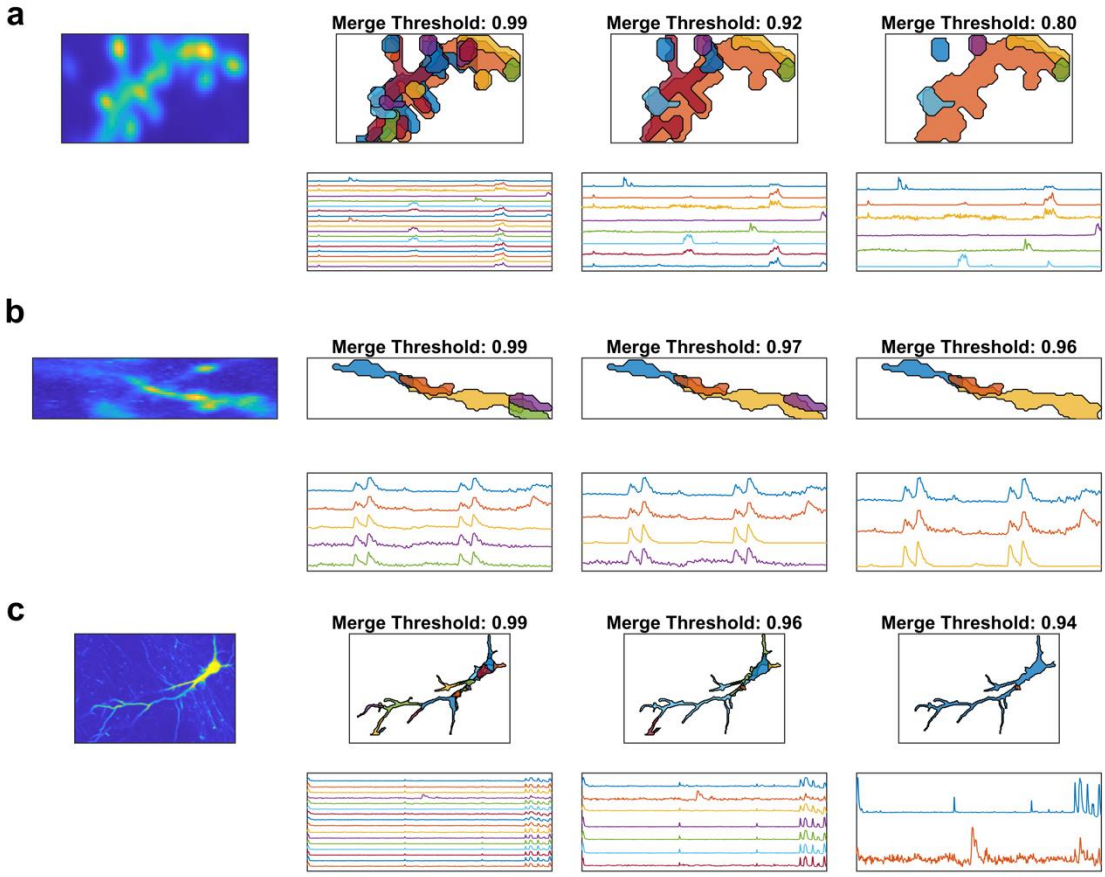
indicated by a sudden drop in minimum within-group correlations (Supplementary Fig. 3, see Methods). Even if users decide not to merge ROIs, correlation-based linkage information can aid in scientific interpretations at later stages."

Discussion (Page 10, Paragraph 1, Lines 415-427):
"As with any data analysis pipeline, the user takes on the responsibility of curating the output of d-NMF to ensure the interpretability and robustness of their results. For instance, depending on the threshold specified for merging ROIs with similar activity, d-NMF can give a very fine or coarse parcellation of the dendritic tree (Supplementary Fig. 10). For those wanting to investigate changes in activity along the dendrite, a high merging threshold would be best, but for those wanting to capture the main global activity across an entire branch, a more permissive merging threshold would be appropriate. Furthermore, the specific optimal threshold may vary across fluorescent indicators, brain region, behavioral state, or species. Additionally, if a dendritic branch is split into multiple ROIs, these will not have independent activity, violating the assumptions of some statistical tests. This can be addressed by calculating summary statistics across ROIs for an entire FOV, or to merge or subsample ROIs with similar activity and verifying that scientific results still hold. Such control analyses should be standard practice when performing population imaging."

Methods (Page 15, Paragraph 5, Lines 673-685):
"We present here a method to estimate a lower bound of a useful merging threshold in Supplementary Fig. 3. This lower bound allows users to avoid linking ROIs from different neurons. Above this lower bound, the choice of merging threshold is up to the user, depending on the scientific question. For a given FOV, all pairwise correlations between the detrended activity traces of ROIs were computed. Then, for a range of threshold values from 1 to 0 (in steps of 0.01), we merged ROIs with correlations above that value. To estimate how well the clustering worked, we identified the cluster with the largest number of member ROIs. We defined the minimum correlation as the lowest correlation value between ROIs in that cluster. This approach is similar to complete-linkage hierarchical clustering, with distance inversely proportional to correlation. We observed that all FOVs had a critical value below which the minimum correlation substantially decreased. This corresponded to ROIs from different neurons becoming linked. We estimate this as the lower bound on an interpretable merging threshold for that FOV."



Supplementary Fig. 3 | Identifying appropriate merging thresholds. a, Left, silhouettes of ROIs belonging to a cluster based on a merging threshold of AAA. Middle, activity traces of the corresponding ROIs, showing extremely high correlation. Right, correlation matrix of the activity traces. **b,** Same as **a** but for a merging threshold of BBB. Note that more ROIs are added to this cluster, but their activity is all highly correlated, indicating they plausibly belong to the same neuron. **c,** Same as **a**, but for a merging threshold of CCC. Note the addition of several ROIs with activity that is uncorrelated with most of the rest of the cluster. **d,** A plot of the minimum correlation value for the largest ROI cluster versus the merge threshold for the FOV in **a-c**. The red circle indicates the threshold at which the merging becomes too permissive (here, 0.65). **e,** Similar threshold-correlation plots for all 8 dense FOVs. The thick line indicates the median across all FOVs. **f,** The median merge threshold across all 8 dense FOVs was 0.69, [0.56, 0.89], n=8 FOVs.



Supplementary Fig. 10 | Exploration of the merging threshold. a, For the dataset in Supplementary Figure 9a, the output of d-NMF is shown for a range of correlation thresholds at which to merge ROIs. The total number of ROIs can vary substantially based on merging threshold, illustrating the importance of choosing the appropriate value for the scientific question at hand. **b,** Same as a but for the dataset in Supplementary Figure 9b. **c,** Same as a but for the dataset in Supplementary Figure 9d.

Q3. I also agree with the reviewer that newer comparisons to Suite2p should be included in the manuscript; and that the the benchmark (F1) is not necessarily a good one. Comparing to human annotators is also not necessarily good since there is arbitrariness here too.

A3. Agreed, we now include the newer comparisons to Suite2p in the revised **Figure 4**.

The rationale for the previous F1 benchmark, computed on individual pixels, was to compare segmentation results when there is ambiguity in how to segment dendritic branches. However, we recognize the reviewers' concerns regarding this approach. As per the reviewers' suggestions, we have come up with a new measure, defining segments based purely on the branching pattern of dendrites, and then quantified the correspondence between those ROIs. This approach is summarized in the **Methods section (Lines 742-769)** and in new **Supplementary Fig. 5b**. Using this new metric, d-NMF retains the highest performance across methods **(Revised Figure 4)**.

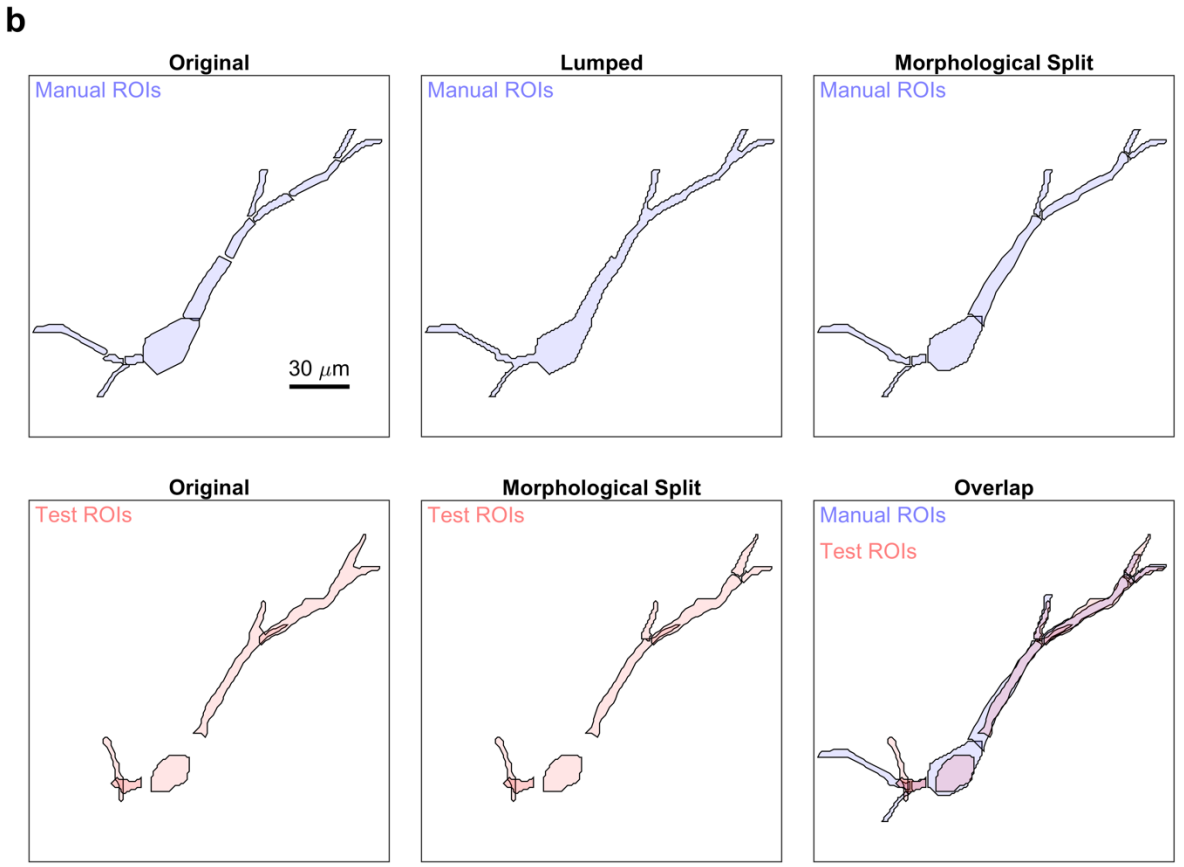
Methods (Page 17, Paragraphs 3-7, Lines 743-770):
"The pixel-wise metric defined above has advantages in its flexibility. It is agnostic to how a neuron may be parceled up, and quantifies completeness of labeling and accuracy of the placement of ROIs. It is most interpretable for sparsely labeled datasets, but has drawbacks when applied to densely labeled datasets. Namely, a single ROI covering all pixels of all neurons in the field of view will achieve the same score as a set of individual ROIs for each dendritic branch. Hence, we defined a second metric based entirely on ROI morphology.

Algorithmically-defined ROIs can often span multiple dendritic branches or encompass soma and proximal branches (Supplementary Fig. 6). To define ROIs in a systematic manner, we performed a morphological split operation on all ROIs (Supplementary Fig. 5), drawing boundaries at branch points and separating putative somatic regions, defined by connected groups of pixels with a radius of at least 10 pixels (11 µm). Resulting ROIs with a size less than 30 pixels were discarded.

Manually-drawn ROIs that sub-segmented a dendritic branch were first merged together and then put through the same morphological split operation. This ensured that manually-drawn ROIs followed the same well-defined guidelines across labelers. For each ROI in set A, we found the ROI in set B that had the highest spatial correspondence, as computed by the Jaccard index (size of the intersection divided by size of the union). We then define a pixel in an ROI as "covered" if it intersects with the matching ROI.

If *A* is the manually-drawn ground truth and *B* is a test set, define:
True Positive (TP) as the number of pixels that are covered in ROIs in set *A*
False Negative (FN) as the number of pixels that are not covered in ROIs in set *A*
False Positive (FP) as the number of pixels that are not covered in ROIs in set *B*

The True Positive Rate (TPR) and F1 score are then computed accordingly. These measures computed ROI-wise are denoted TPR_{roi} and $F1_{roi}$, respectively."



Supplementary Fig. 5 | Evaluating ROI segmentation performance. b, Demonstration of ROI evaluation in the real datasets. Top, the original ROIs drawn by human annotators are grouped by cell identity to make a "Lumped" ROI, which is then morphologically split at branch points. Bottom, algorithmically defined ROIs are first morphologically split, then the best matches to manual ROIs are identified. Pixels covered by both a manual ROI and test ROI (blue and red) are counted as True positives. Pixels only in manual ROIs (blue) are False Negatives. Pixels only in Test ROIs (red) are False Positives.

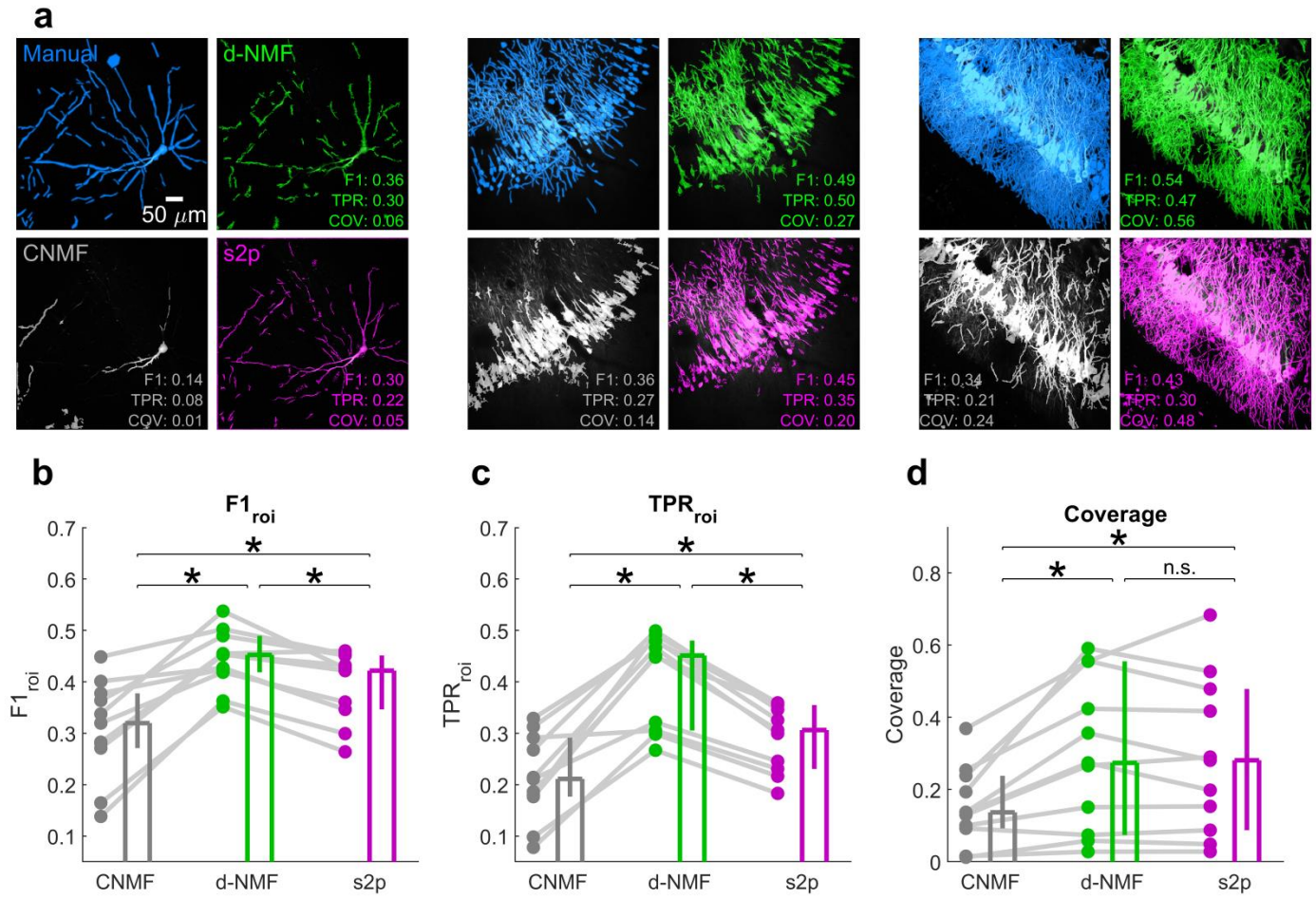


Fig. 4 | d-NMF accurately labels ROIS with highest coverage. b, The $F1_{roi}$ score of identified ROIs (compared to manual labeling) for d-NMF (green, 0.45, [0.42, 0.49]) was significantly higher (p=9.8x10-4) than that of CNMF (gray, 0.32, [0.27, 0.38]), and significantly higher (p=0.002) than suite2p (s2p, purple: 0.42, [0.35, 0.45]). The F1 score of suite2p was also higher than CNMF (p=0.014). c, The True Positive Rate (TPR_{roi}) of identified ROIs (compared to manual labeling) for d-NMF (0.45, [0.31, 0.48]) was significantly higher (p=9.8x10-4) than that of CNMF (0.21, [0.18, 0.29]), and significantly higher (p=9.8x10-4) than suite2p (0.31, [0.23, 0.35]), which was also higher than CNMF (p=0.0068). d, d-NMF had significantly higher overall coverage than CNMF, but not suite2p (d-NMF Coverage: 0.27, [0.07, 0.55]; CNMF Coverage: 0.14, [0.09, 0.24]; suite2p Coverage: 0.28, [0.09, 0.48]; d-NMF vs CNMF: p=0.0029; d-NMF vs suite2p: p=0.52; CNMF vs suite2p: p=0.002). n = 11 FOVs, two-sided Wilcoxon signed-rank test for all. Values reported as median and 95% confidence interval of the median.

Using human annotations is a standard in the field, but we recognize the points raised by the reviewers regarding limitations to this approach, which we state in our **Methods (Page 17, Paragraph 1, Lines 738-741)**:

"This measure is a combination of recall (True Positive Rate, TP/(TP+FN)) and precision (Positive Predictive Value, TP/(TP+FP)), as human-labeled datasets may have incomplete labeling, which would yield poor estimates of false positives."

Q4. Here again, the authors could address these issues relatively easily with synthetic data.

A4. Indeed, synthetic data is a good solution for these issues. We did utilize this approach in **Figure 3**, where we generate synthetic data with known ground truth for ROIs and their activity, and show that d-NMF is more accurate even at very high densities.

Reviewer #2 (Remarks to the Author):

*Q5-1. In this rebuttal, Moore et al clarified a few things for me. While some things got better, others did not change and others got worse. For my main criticisms in the previous rounds of review, the authors essentially answered by saying that dealing with the bAP and with the oversplitting of the dendrites is the user's problem. They illustrate how one might do that, which I appreciate. For example, they show individual examples of local and global activity and they show me "privately" a coarse-to-fine segmentation with an off-the-shelf hierarchical merging algorithm (but they leave this figure out of the manuscript). I am not sure what to do with this response. My comment was simply that I consider these two steps as *the* problem that needs to be solved for dendritic segmentation. Whether one gets 60% or 70% pixels covered by the dendrite segments is of little concern to the user. This paper is essentially trying to argue that the segments found by their algorithm are of a better quality in some sense than the ones that could have been found with non-dedicated algorithms. But for a full analysis, the users still need to merge the segments themselves, which can then be used to estimate global vs local activity. This merging step is crucial, not because we only want a single activity number for each neuron as the authors imply that I meant, but because we need to know which segments and which branches go together, so we can do the kinds of studies the authors enumerated in their reviewer table 1 which they made just for me. While the hierarchical merging algorithm is a start, it's a very basic approach, and there is no quantification of how well it works. I thus consider that the status of my two main previous concerns is unchanged.*

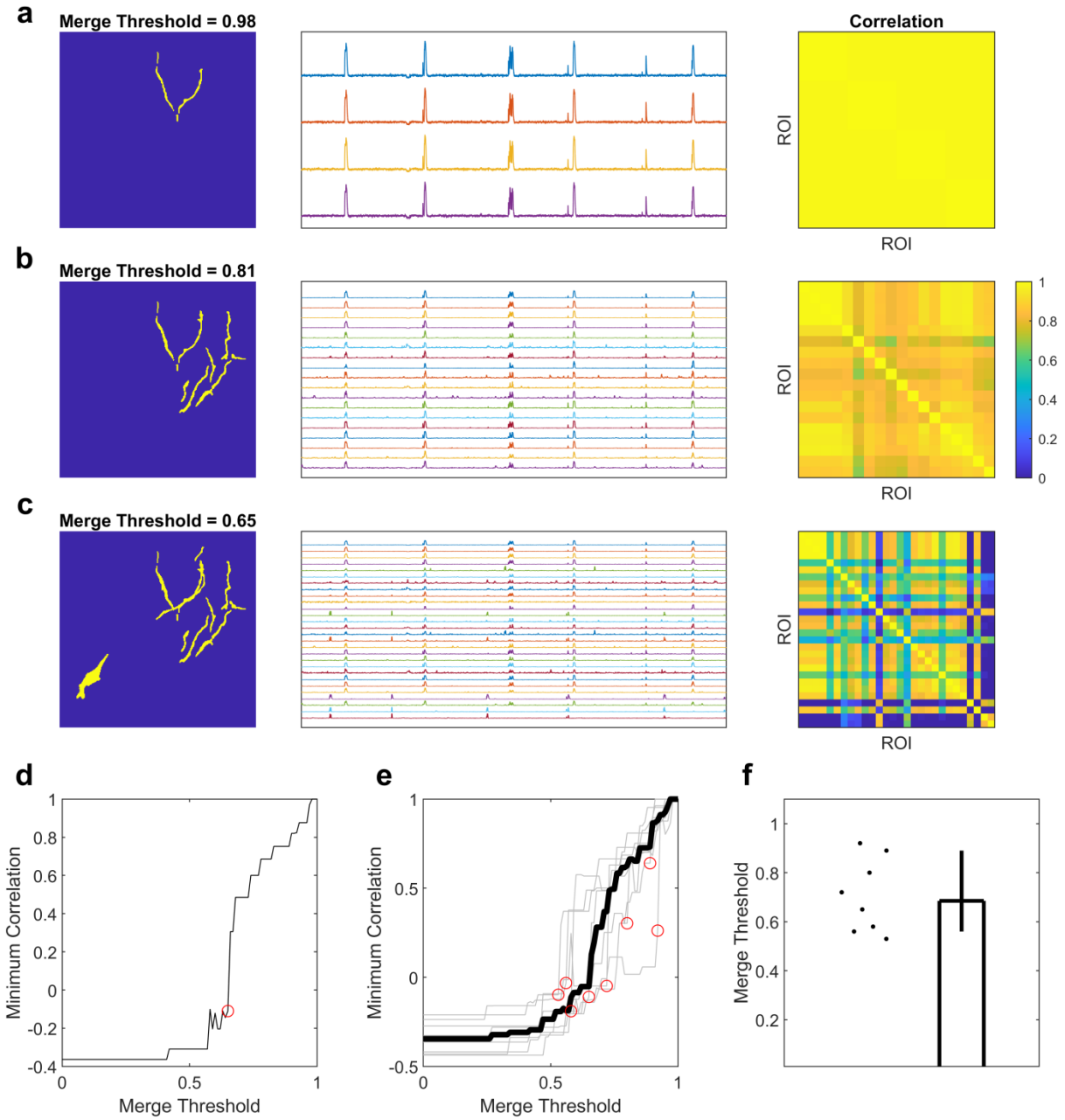
A5-1. In agreement with the other reviewers, we believe that choosing the best analytical approach based on experimental goals and data quality and related interpretations has to be at the user's discretion. However, based on the reviewer's suggestion, we now include the coarse-to-fine segmentation analyses which can serve as a guide to users **(Supplementary Figs. 3, 10)**. We also include sections in the Results **(Lines 153-162)**, **Discussion (Lines 415-427),** and **Methods (Lines 673-685)** emphasizing that users must take an active approach in curating the output of our algorithm.

Results (Page 4, Paragraph2, Lines 153-162):
"In the Merging step, we combine ROIs (Supplementary Table 2c) that are likely to correspond to the same functional unit. Long dendritic branches may span multiple image patches (Fig. 2a) or may be represented by multiple ROIs due to local variations in activity. The algorithm offers users the ability to set a threshold correlation with which to merge ROIs. A high threshold can result in a fine parcellation of dendritic structure, preserving local differences in activity, while a low threshold will merge ROIs with broadly the same activity. We observed that a general rule on the lower bound of the merging threshold is to identify when ROIs not belonging to the same neuron become linked together. This is indicated by a sudden drop in minimum within-group correlations (Supplementary Fig. 3, see Methods). Even if users decide not to merge ROIs, correlation-based linkage information can aid in scientific interpretations at later stages."

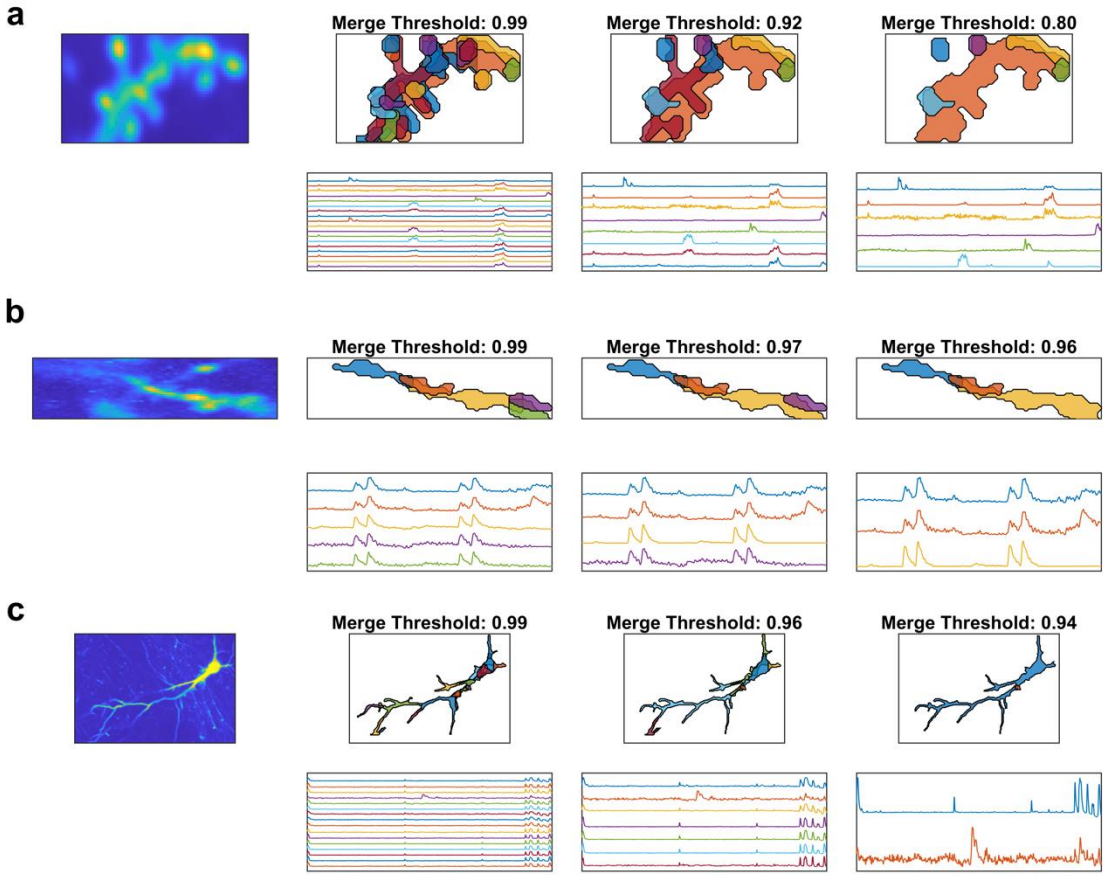
Discussion (Page 10, Paragraph 1, Lines 415-427):
"As with any data analysis pipeline, the user takes on the responsibility of curating the output of d-NMF to ensure the interpretability and robustness of their results. For instance, depending on the threshold specified for merging ROIs with similar activity, d-NMF can give a very fine or coarse parcellation of the dendritic tree (Supplementary Fig. 10). For those wanting to investigate changes in activity along the dendrite, a high merging threshold would be best, but for those wanting to capture the main global activity across an entire branch, a more permissive merging threshold would be appropriate. Furthermore, the specific optimal threshold may vary across fluorescent indicators, brain region, behavioral state, or species. Additionally, if a dendritic branch is split into multiple ROIs, these will not have independent activity, violating the assumptions of some statistical tests. This can be addressed by calculating summary statistics across ROIs for an entire FOV, or to merge or subsample ROIs with similar activity and verifying that scientific results still hold. Such control analyses should be standard practice when performing population imaging."

Methods (Page 15, Paragraph 5, Lines 673-685):
"We present here a method to estimate a lower bound of a useful merging threshold in Supplementary Fig. 3. This lower bound allows users to avoid linking ROIs from different neurons. Above this lower bound, the choice of merging threshold is up to the user, depending on the scientific question. For a given FOV, all pairwise correlations between the detrended activity traces of ROIs were computed. Then, for a range of threshold values from 1 to 0 (in steps of 0.01), we merged ROIs with correlations above that value. To estimate how well the clustering worked, we identified the cluster

with the largest number of member ROIs. We defined the minimum correlation as the lowest correlation value between ROIs in that cluster. This approach is similar to complete-linkage hierarchical clustering, with distance inversely proportional to correlation. We observed that all FOVs had a critical value below which the minimum correlation substantially decreased. This corresponded to ROIs from different neurons becoming linked. We estimate this as the lower bound on an interpretable merging threshold for that FOV."



Supplementary Fig. 3 | Identifying appropriate merging thresholds. a, Left, silhouettes of ROIs belonging to a cluster based on a merging threshold of AAA. Middle, activity traces of the corresponding ROIs, showing extremely high correlation. Right, correlation matrix of the activity traces. **b,** Same as **a** but for a merging threshold of BBB. Note that more ROIs are added to this cluster, but their activity is all highly correlated, indicating they plausibly belong to the same neuron. **c,** Same as **a**, but for a merging threshold of CCC. Note the addition of several ROIs with activity that is uncorrelated with most of the rest of the cluster. **d,** A plot of the minimum correlation value for the largest ROI cluster versus the merge threshold for the FOV in **a-c**. The red circle indicates the threshold at which the merging becomes too permissive (here, 0.65). **e,** Similar threshold-correlation plots for all 8 dense FOVs. The thick line indicates the median across all FOVs. **f,** The median merge threshold across all 8 dense FOVs was 0.69, [0.56, 0.89], n=8 FOVs.



Supplementary Fig. 10 | Exploration of the merging threshold. a, For the dataset in Supplementary Figure 9a, the output of d-NMF is shown for a range of correlation thresholds at which to merge ROIs. The total number of ROIs can vary substantially based on merging threshold, illustrating the importance of choosing the appropriate value for the scientific question at hand. **b,** Same as a but for the dataset in Supplementary Figure 9b. **c,** Same as a but for the dataset in Supplementary Figure 9d.

A5-2. There may be some ambiguity in the interpretation of the term "merging," which has led to some misunderstanding. Our use of the term "merging" means combining smaller ROIs into larger ROIs, which would discard the differences in activity in those smaller ROIs. On the other hand, "linking" ROIs that belong to the same neuron is indeed useful to be able to relate the activity of dendrites with their parent soma, which we explore in the new **Supplementary Figure 3**. But we believe that a satisfactory characterization of the strengths and drawbacks of different merging or linking approaches would require significant space, effort, and even novel datasets with structural indicators, and that the tool and scientific results in this manuscript stand on their own. We provide one way to select merging threshold and guidance for the users on how to do so – as described above.

Reviewer table 1 was meant as a broad overview to emphasize how useful insights into dendritic function were made in other studies without dissociating bAPs as well as highlighting how, depending on the goal of the study, one may need to define small ROI segments; we cite these studies in the manuscript, and now include a dedicated section bringing emphasizing these discussion points, in **Lines 488-499**.

Discussion (Page 11, Paragraph 2, Lines 488-499):
"Many valuable insights about dendritic function have been gleaned without explicitly delineating local dendritic events from back-propagating events, across different brain areas such as motor cortex[22,25,79], visual cortex[1,23,80,81], hippocampal CA1[21,33,82], and lateral amygdala[83]. Definitively resolving back-propagating action potentials from local dendritic activity

requires deliberately experimental design involving technically challenging approaches[37,84]. For accurately dissociating the source of dendritic activity (forward vs. back propagating) one needs to establish ground truth with simultaneous dendritic and somatic electrophysiology and imaging coupled with well-defined stimulation protocols[85,86]. Given the slower kinetics and lower single spike resolution of Ca^{2+} indicators, resolving bAPs requires further tool development such as in vivo optimized high signal-to-noise ratio voltage imaging sensors[87]. The analysis tools presented here will synergize with future developments in these experimental areas to enable principled investigation into dendritic signaling."

As tools and assurances to the user, we have demonstrated that our algorithm does not systematically oversplit compared to other methods **(Supplementary Fig. 6)**, and provided a way to merge or link ROIs after processing **(Supplementary Figs. 3, 10)**.

Q6. Now for the things that got better: the authors showed that their scientific analyses do not depend on the level of merging, mainly because the analyses are done on an FOV basis. Further, they tell me (but not the reader), that this level of control analyses should be considered standard going further. Are you sure the users will never do statistics at the level of ROIs? Sometimes this is unavoidable scientifically, and also often I suspect the users will not know that this is not an OK thing to do.

A6. We are glad that we were able to clarify this point. With regard to how users may interpret our output, we must trust that users will follow best scientific practices. We include this point emphasizing best practices in the revised discussion section **(Page 10, Paragraph 1, Lines 415-427)**:

"As with any data analysis pipeline, the user takes on the responsibility of curating the output of d-NMF to ensure the interpretability and robustness of their results. For instance, depending on the threshold specified for merging ROIs with similar activity, d-NMF can give a very fine or coarse parcellation of the dendritic tree (Supplementary Fig. 10). For those wanting to investigate changes in activity along the dendrite, a high merging threshold would be best, but for those wanting to capture the main global activity across an entire branch, a more permissive merging threshold would be appropriate. Furthermore, the specific optimal threshold may vary across fluorescent indicators, brain region, behavioral state, or species. Additionally, if a dendritic branch is split into multiple ROIs, these will not have independent activity, violating the assumptions of some statistical tests. This can be addressed by calculating summary statistics across ROIs for an entire FOV, or to merge or subsample ROIs with similar activity and verifying that scientific results still hold. Such control analyses should be standard practice when performing population imaging."

Q7. Now for the things that got worse, and only got worse because I realized the true nature of the benchmarks performed here. The authors evaluate this ROI segmentation problem as if it is a semantic segmentation task. This is inappropriate, and tells us very little about how well any of the algorithms work. An algorithm would get the same score if all the ROIs are split into literally individual pixels, or if they are all combined into a single ROI for the entire field of view. I think the reason I missed this originally is that I could not conceive of making this choice for evaluating a clear instance segmentation problem. There are so many ways to do well on a semantic segmentation task, none of which are shown (like thresholding a locally-normalized intensity map or maximum projection, or a correlation map, or a skewness map, or a variance map etc). But that wouldn't really capture at all the thing we care about, which is that the segments should correspond to meaningful, individually separated parts of neurons. In light of the previous rounds of reviews, no benchmark would really capture this, since the human annotators essentially split these dendrites into arbitrary segments, just making sure they are not too long.

A7. We will first clarify that human annotation was not arbitrary, as we state that "dendritic branches were split up into multiple ROIs at branch points" in **Lines 711-712**. This was done by expert neuroscientists who are cognizant of biological morphology and functional compartmentalization of neurons. This lends itself to a segmentation that has

meaningful, individually separated parts of neurons. The point of contention is the subsegmenting of long ROIs; otherwise, the manual annotations are by no means arbitrary.

We recognize that the pixel-wise measure is more appropriate for sparsely-labeled datasets. As per the reviewers' suggestions, we have come up with a new measure, defining segments based purely on the branching pattern of dendrites, and then quantified the correspondence between those ROIs. This approach is summarized in the Methods section **(Lines 743-770)** and in new **Supplementary Fig. 5b**. The metrics reported in **Fig. 4** and **Supplementary Fig. 12** have been updated to report this new measure. Using this new metric, d-NMF retains the highest performance across methods (Revised **Figure 4**).

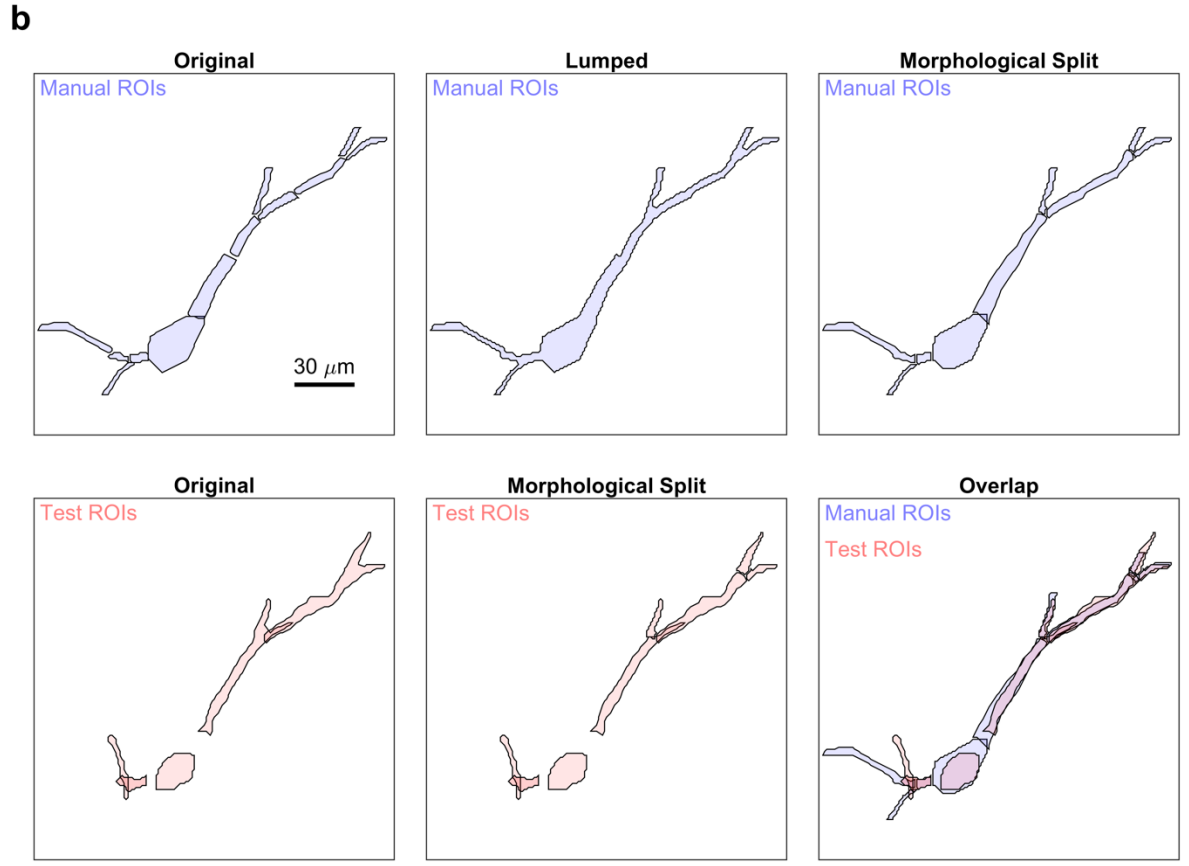
Methods (Page 17, Paragraphs 3-7, Lines 743-770):
"The pixel-wise metric defined above has advantages in its flexibility. It is agnostic to how a neuron may be parceled up, and quantifies completeness of labeling and accuracy of the placement of ROIs. It is most interpretable for sparsely labeled datasets, but has drawbacks when applied to densely labeled datasets. Namely, a single ROI covering all pixels of all neurons in the field of view will achieve the same score as a set of individual ROIs for each dendritic branch. Hence, we defined a second metric based entirely on ROI morphology.

Algorithmically-defined ROIs can often span multiple dendritic branches or encompass soma and proximal branches (Supplementary Fig. 6). To define ROIs in a systematic manner, we performed a morphological split operation on all ROIs (Supplementary Fig. 5), drawing boundaries at branch points and separating putative somatic regions, defined by connected groups of pixels with a radius of at least 10 pixels (11 µm). Resulting ROIs with a size less than 30 pixels were discarded.

Manually-drawn ROIs that sub-segmented a dendritic branch were first merged together and then put through the same morphological split operation. This ensured that manually-drawn ROIs followed the same well-defined guidelines across labelers. For each ROI in set A, we found the ROI in set B that had the highest spatial correspondence, as computed by the Jaccard index (size of the intersection divided by size of the union). We then define a pixel in an ROI as "covered" if it intersects with the matching ROI.

If *A* is the manually-drawn ground truth and *B* is a test set, define:
True Positive (TP) as the number of pixels that are covered in ROIs in set *A*
False Negative (FN) as the number of pixels that are not covered in ROIs in set *A*
False Positive (FP) as the number of pixels that are not covered in ROIs in set *B*

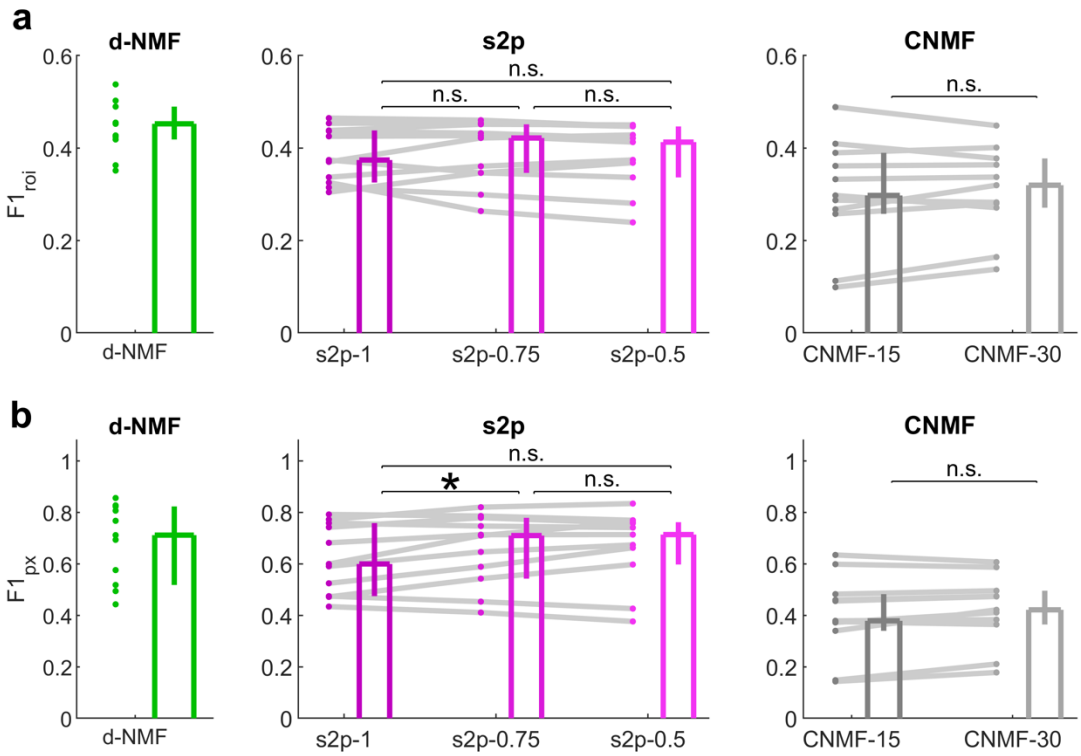
The True Positive Rate (TPR) and F1 score are then computed accordingly. These measures computed ROI-wise are denoted TPR_{roi} and $F1_{roi}$, respectively."



Supplementary Fig. 5 | Evaluating ROI segmentation performance. b, Demonstration of ROI evaluation in the real datasets. Top, the original ROIs drawn by human annotators are grouped by cell identity to make a "Lumped" ROI, which is then morphologically split at branch points. Bottom, algorithmically defined ROIs are first morphologically split, then the best matches to manual ROIs are identified. Pixels covered by both a manual ROI and test ROI (blue and red) are counted as True positives. Pixels only in manual ROIs (blue) are False Negatives. Pixels only in Test ROIs (red) are False Positives.

Q8. I also noticed that Suite2p did in fact "improve" to an F1 score of 0.68 with the threshold of 0.75, which was in fact the same as d-NMF. However, this figure was only shown to me, and not included in the paper. Suite2p-0.75 should become the default comparison to use in the main text. Interestingly, a difference between 0.59 and 0.68 was reported as significant between d-NMF and Suite2p in the main figure, but the same mean values came out insignificant when varying the threshold of Suite2p for my "private" figure. This is probably not a meaningful distinction. I don't think these scores are very meaningful in general, but I also don't understand why the authors would hide the fact that simply decreasing a threshold results in more pixels in the output, and that's the main determinant of performance by this metric.

A8. Including reviewer figures is a common practice. We did so considering the high number of supplemental figures (15 at the last point of review) without an intention to hide results, as the contents of these reviews are publicly accessible. The new **Supplementary Fig. 7** now contains this information using both the pixel-wise evaluation criterion and the ROI-wise measure. We now exclusively compare to Suite2p-0.75 **(Fig. 4)** using the new metric we describe in point 7 above.



c

Measure	d-NMF	s2p-1	s2p-0.75	s2p-0.5	CNMF-15	CNMF-30
$F1_{roi}$	0.45	0.37	0.42	0.41	0.30	0.32
$F1_{px}$	0.71	0.60	0.71	0.71	0.38	0.42
# ROIs	1190	812	1507	2165	109	164
ROI Size (μm^2)	155.0	108.3	92.0	86.1	473.5	450.2
ROI Length (μm)	37.2	32.1	30.3	25.5	53.4	50.7

Supplementary Fig. 7 | Adjusting thresholds for segmentation methods. a, Left, for reference, the $F1_{roi}$ score for d-NMF (0.45, [0.42, 0.49], *n* = 11 FOVs for all). Middle, suite2p had a small change in $F1_{roi}$ score when the "threshold" parameter was lowered from 1 (0.37, [0.33, 0.44]) to 0.75 (0.42, [0.35, 0.45]) to 0.5 (0.41, 0.34, 0.45]); s2p-1 vs s2p-0.75, p=0.97; s2p-1 vs s2p-0.5, p=0.70; s2p-0.75 vs s2p-0.5, p=0.24, two-sided Wilcoxon signed-rank test). Right, CNMF output was slightly more accurate when the "number of clusters" parameter was changed (CNMF-15, 0.30, [0.26, 0.39]; CNMF-30, 0.32, [0.27, 0.38]; p=0.46, two-sided Wilcoxon signed-rank test). d-NMF vs s2p-0.75: $p=2.0x10^{-3}$; d-NMF vs CNMF-30: $p=9.8x10^{-4}$; s2p-0.75 vs CNMF-30: p=0.014, two-sided Wilcoxon signed-rank test for all. **b,** Same as **a** but for the $F1_{px}$ score. d-NMF: 0.71, [0.52, 0.82]. s2p-1: 0.60, [0.47, 0.76]; s2p-0.75: 0.71, [0.54, 0.78]; s2p-0.5: 0.71, [0.60, 0.76]; s2p-1 vs s2p-0.75, p=0.042; s2p-1 vs s2p-0.5, p=0.15; s2p-0.75 vs s2p-0.5, p=0.52, two-sided Wilcoxon signed-rank test; CNMF-15, 0.38, [0.34, 0.48]; CNMF-30, 0.42, [0.36, 0.50]; p=0.067, two-sided Wilcoxon signed-rank test. d-NMF vs s2p-0.75: p=0.083; d-NMF vs CNMF-30: $p=9.8x10^{-4}$; s2p-0.75 vs CNMF-30: $p=9.8x10^{-4}$, two-sided Wilcoxon signed-rank test for all. **c,** Table of median values of $F1_{roi}$ score, $F1_{px}$ score, number of ROIs, ROI Size, and ROI Length for easy comparison.

Q9. It is at the discretion of the editor what to do with these comments, but I am afraid I cannot give a more positive review. I am sorry I missed some of these crucial details on the initial review, and I thank the authors for pointing them out in the rebuttal.

A9. We are glad we could resolve these issues as a result of these discussions.