github.com/con/opfvta-reexecution

Neuroimaging Article Reexecution and Reproduction Assesment System

Horea-Ioan Ioanas[1], Austin Macdonald[1], Yaroslav O. Halchenko[1]

[1]Center for Open Neuroscience, Department of Psychological and Brain Sciences, Dartmouth College

Abstract

The value of research articles is increasingly contingent on the results of complex data analyses which substantiate their claims. Compared to data production, data analysis more readily lends itself to a higher standard of both full transparency and repeated operator-independent execution. This higher standard can be approached via fully reexecutable research outputs, which contain the entire instruction set for end-to-end generation of an entire article solely from the earliest feasible provenance point, in a programatically executable format. In this study, we make use of a peer-reviewed neuroimaging article which provides complete but fragile reexecution instructions, as a starting point to formulate a new reexecution system which is both robust and portable. We render this system modular as a core design aspect, so that reexecutable article code, data, and environment specifications could potentially be substituted or adapted. In conjunction with this system, which forms the demonstrative product of this study, we detail the core challenges with full article reexecution and specify a number of best practices which permitted us to mitigate them. We further show how the capabilities of our system can subsequently be used to provide reproducibility assessments, both via simple statistical metrics and by visually highlighting divergent elements for human inspection. We argue that fully reexecutable articles are thus a feasible best practice, which can greatly enhance the understanding of data analysis variability and the trust in results. Lastly, we comment at length on the outlook for reexecutable research outputs and encourage re-use and derivation of the system produced herein.

Workflow

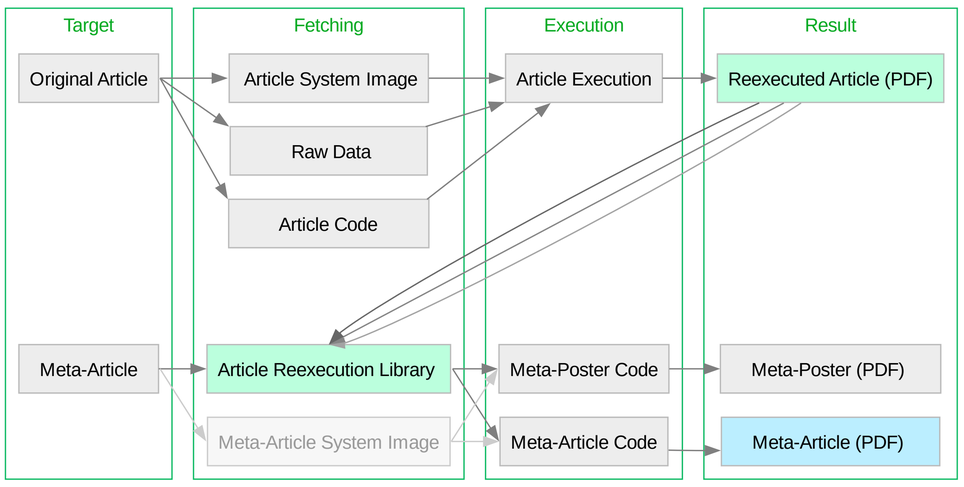


Figure 1: The reexecution system encompasses both the original article (first target), and the "meta-article" publishing materials (article manuscript, as well as this poster), the latter of which takes user- and developer-submitted reexecution results as an input for the reproduction quality assessment.

Topoplogy

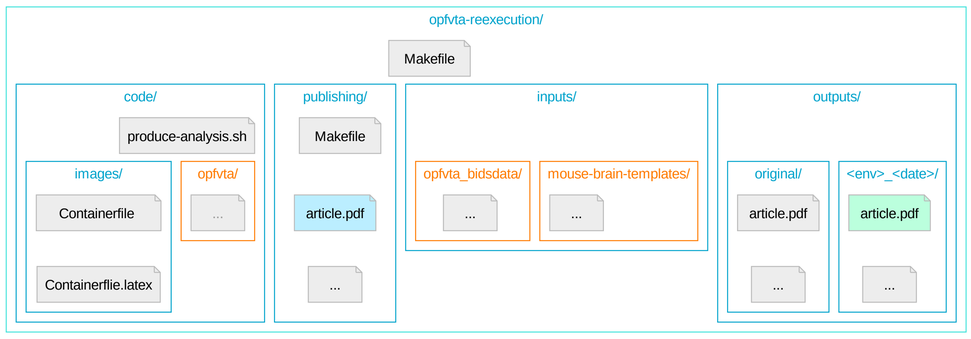


Figure 2: The reexecution workflow is supported by a resource topology in which reexecution code (first box), "meta-article" code (second box), reexecution resources (third box), and the reexecution output record (last box) are separated at the top directory level. The figure depicts direcotry trees via nested boxes, with external resources automatically fetched as via the reexecution code being highlighted in orange. The green highlighted article represents a sample reexecution output, and the blue hignlighted article represents the manuscript, an analogous output to this poster generated in the same directory.

Best Practice Guidelines

As part of setting up an encompassing reexecution system, we formulate a number of best practices, including:

- **Errors should be fatal more often than not.**
  `set -eu`, prepended to POSIX shell scripts, will ensure that workflows fail when a subcommand does, or when an encountered variable is undefined.
- **Avoid assuming a directory context for execution.**
  `cd "$(dirname "$0")"`, prepended to POSIX shell scripts, will ensure that in complex workflows scripts can operate relative to their location directory context and not the execution context.
- **Workflow granularity greatly benefits efficiency.**
  While the underlying execution system of the target article, RepSeP [1] separates data analysis into two distinct (voxel-space "low iteration" and top-level "high iteration") steps, further granularity can benefit debugging, particularly in container environments.
- **Resources should be bundled into a DataLad superdataset.**
  Resource bundling, with usage of submodules for external resources (as seen in fig. 2) allows management of required resources via Git and associated technologies, such as DataLad [2] — this is known as the YODA principle [3].
- **Dependency versions inside container environments should be frozen as soon as feasible.**
  This is best accomplished via a package manager which uses version tracking for its software provision index; in Gentoo Linux, used here on account of broad provision of neuroscience packages [4], this can be done via:
  `cd /.../myrepo; git fetch origin $myhash; git checkout $myhash`.

References

[1] H.-I. Ioanas and M. Rudin, "Reproducible self-publishing for Python-based research," EuroSciPy, Aug. 2018.

[2] Y. Halchenko, K. Meyer, B. Poldrack, D. Solanky, A. Wagner, J. Gors, D. MacFarlane, D. Pustina, V. Sochat, S. Ghosh, C. Mönch, C. Markiewicz, L. Waite, I. Shlyakhter, A. de la Vega, S. Hayashi, C. Häusler, J.-B. Poline, T. Kadelka, K. Skytén, D. Jarecka, D. Kennedy, T. Strauss, M. Cieslak, P. Vavra, H.-I. Ioanas, R. Schneider, M. Pflüger, J. Haxby, S. Eickhoff, and M. Hanke, *DataLad: distributed system for joint management of code, data, and their relationship*, vol. 6. The Open Journal, July 2021.

[3] M. Hanke, M. Visconti di Oleggio Castello, K. Meyer, B. Poldrack, and Y. O. Halchenko, "YODA: YODA's organigram on data analysis." Poster presented at the annual meeting of the Organization for Human Brain Mapping, Singapore, 2018.

Reproduction Assessment Showcase

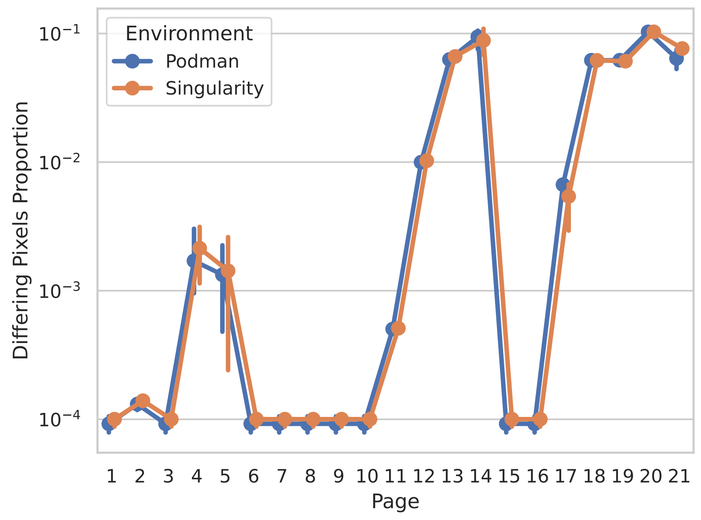


Figure 3: Page-wise pixel difference comparison across multiple reexecutions in different environments indicates consistency of variability in both extent and location.

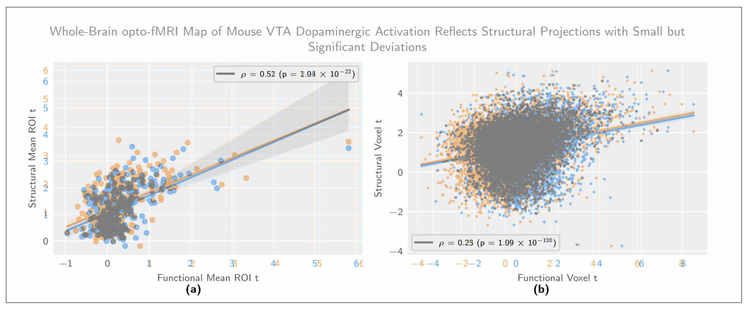
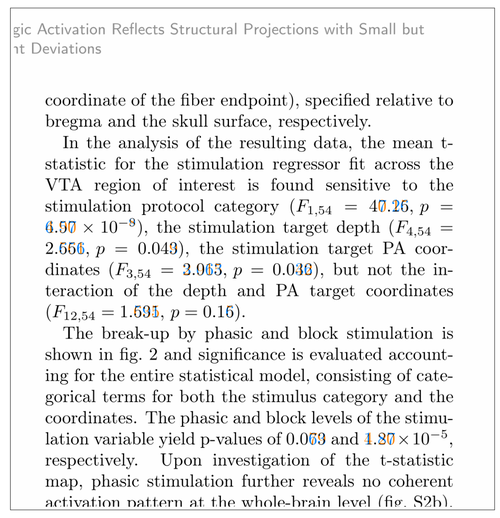


Figure 4: One notable source of variability are data plots, where it can be observed that even as data points vary to an almost full extent, statistical summaries can remain constant.


gic Activation Reflects Structural Projections with Small but
nt Deviations

coordinate of the fiber endpoint), specified relative to bregma and the skull surface, respectively.

In the analysis of the resulting data, the mean t-statistic for the stimulation regressor fit across the VTA region of interest is found sensitive to the stimulation protocol category ($F_{1,54} = 47.26$, $p = 6.50 \times 10^{-9}$), the stimulation target depth ($F_{4,54} = 2.666$, $p = 0.043$), the stimulation target PA coordinates ($F_{3,54} = 3.063$, $p = 0.036$), but not the interaction of the depth and PA target coordinates ($F_{12,54} = 1.595$, $p = 0.16$).

The break-up by phasic and block stimulation is shown in fig. 2 and significance is evaluated accounting for the entire statistical model, consisting of categorical terms for both the stimulus category and the coordinates. The phasic and block levels of the stimulation variable yield p-values of 0.063 and 4.30×10^{-5}, respectively. Upon investigation of the t-statistic map, phasic stimulation further reveals no coherent activation pattern at the whole-brain level (fig. S2b).

The main and interaction effects of the implant coordinate variables are better described categorically than linearly (figs. S1 and 2b). Consequently, the most suitable implant coordinate group for the assay can best be determined on the basis of categorical classification of implant coordinates. We classify the implant coordinates into a "best" and a "rejected" group by k-means clustering the aggregate VTA t-statistic scores into two clusters, and find spatial coherence for the "best" coordinate group (categorization highlighted in fig. 2b).

For block stimulation, the best implant category group (fig. 3a) and the rejected implant category group (fig. 3c) show not only a difference in overall stimulus-evoked signal intensity, but also a difference in efferent distribution, with the rejected implant category efferent spectrum more strongly weighted towards caudal brain areas. This distinction specifically arises for implant categorization based on block scan VTA t-statistic means, and is not as salient if implants are categorized based on a posteroanterior implant coordinate delimiter (fig. S3).

The activation pattern elicited by block stimulation in the best implant category group shows strong coherent clusters of activation. The top activation areas are predominantly located in the right hemisphere, with highly significant laterality ($p = 8.37 \times 10^{-5}$) seen in the comparison of left and right hemisphere atlas parcellation region averages. Activation is seen in regions surrounding the stimulation site, such as the ventral tegmental decussation and the interpeduncular nucleus. The largest activation cluster encompasses well-known dopaminergic VTA projection areas in the subcortical rostroventral regions of the brain (nucleus accumbens, striatum, and the basal forebrain), with weaker activation observed in smaller structures in the vicinity of these regions, such as the fasciculus retroflexus, anterior commissure and the claustrum.

2023-07-25

[1], $\approx 10,000$ in rats [2],
tography commonly fails
ty of this neurotransmit-
n being a prominent node

tional connectivity betwe
and associated projectio

Key questions surrou
clinical models are, firstl

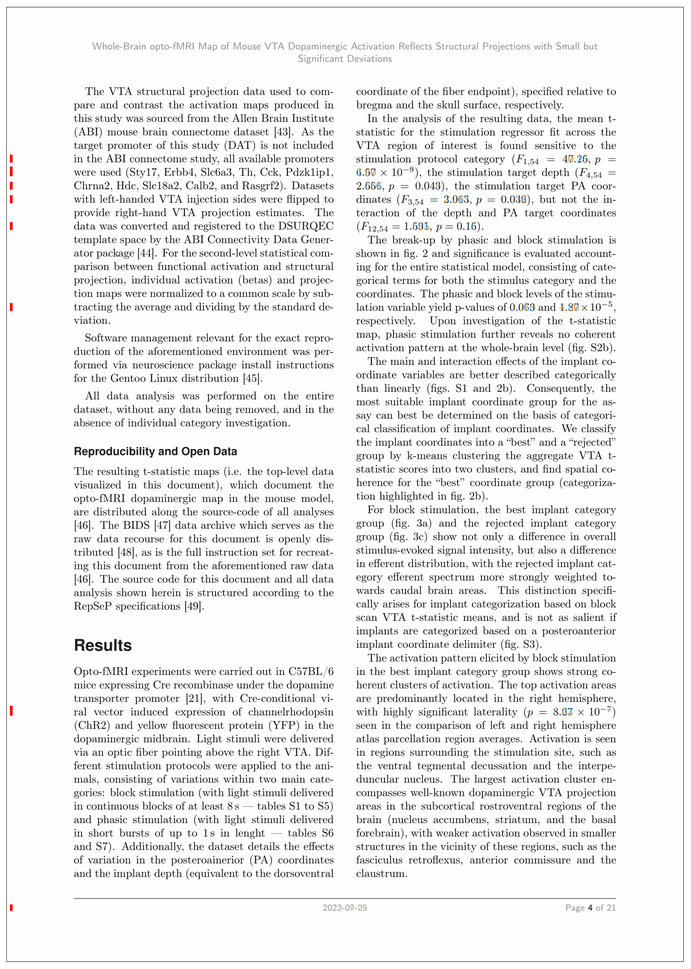
2023-07-25

Figure 5: Text differences in statistical summaries account for a small proportion of pixel differences, but can remain well-localized instead of spreading via test shift if statistical summaries are appropriately trimmed down to a constant length.

Figure 6: A good litmus test for monitoring differences (accounting for the baseline difference in fig. 3) is the datestamp of the reexecution, which should always be expected to differ from the manuscript.

Full Document Comparison

Reproduction assessment is based on *full* document "diffs". The following figures are excerpts, with tinted highlighting (blue for the original manuscript, and orange for reexecution result). First row pages exemplify inline statistical differences and second row pages exemplify figure differences. Differing sections are highlighted with a red left-hand marking.


Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

The VTA structural projection data used to compare and contrast the activation maps produced in this study was sourced from the Allen Brain Institute (ABI) mouse brain connectome dataset [43]. As the target promoter of this study (DAT) is not included in the ABI connectome study, all available promoters were used (Sty17, Erbb4, Slc6a3, Th, Cck, Pdzk1ip1, Chrna2, Hdc, Slc18a2, Calb2, and Rasgrf2). Datasets with left-handed VTA injection sides were flipped to provide right-hand VTA projection estimates. The data was converted and registered to the DSURQEC template space by the ABI Connectivity Data Generator package [44]. For the second-level statistical comparison between functional activation and structural projection, individual activation (betas) and projection maps were normalized to a common scale by subtracting the average and dividing by the standard deviation.

Software management relevant for the exact reproduction of the aforementioned environment was performed via neuroscience package install instructions for the Gentoo Linux distribution [45].

All data analysis was performed on the entire dataset, without any data being removed, and in the absence of individual category investigation.

Reproducibility and Open Data

The resulting t-statistic maps (i.e. the top-level data visualized in this document), which document the opto-fMRI dopaminergic map in the mouse model, are distributed along the source-code of all analyses [46]. The BIDS [47] data archive which serves as the raw data recourse for this document is openly distributed [48], as is the full instruction set for recreating this document from the aforementioned raw data [46]. The source code for this document and all data analysis shown herein is structured according to the RepSeP specifications [49].

Results

Opto-fMRI experiments were carried out in C57BL/6 mice expressing Cre recombinase under the dopamine transporter promoter [21], with Cre-conditional viral vector induced expression of channelrhodopsin (ChR2) and yellow fluorescent protein (YFP) in the dopaminergic midbrain. Light stimuli were delivered via an optic fiber pointing above the right VTA. Different stimulation protocols were applied to the animals, consisting of variations within two main categories: block stimulation (with light stimuli delivered in continuous blocks of at least 8 s — tables S1 to S5) and phasic stimulation (with light stimuli delivered in short bursts of up to 1 s in lenght — tables S6 and S7). Additionally, the dataset details the effects of variation in the posteroainerior (PA) coordinates and the implant depth (equivalent to the dorsoventral coordinate of the fiber endpoint), specified relative to bregma and the skull surface, respectively.

In the analysis of the resulting data, the mean t-statistic for the stimulation regressor fit across the VTA region of interest is found sensitive to the stimulation protocol category ($F_{1,54} = 47.26$, $p = 6.50 \times 10^{-9}$), the stimulation target depth ($F_{4,54} = 2.666$, $p = 0.043$), the stimulation target PA coordinates ($F_{3,54} = 3.063$, $p = 0.036$), but not the interaction of the depth and PA target coordinates ($F_{12,54} = 1.595$, $p = 0.16$).

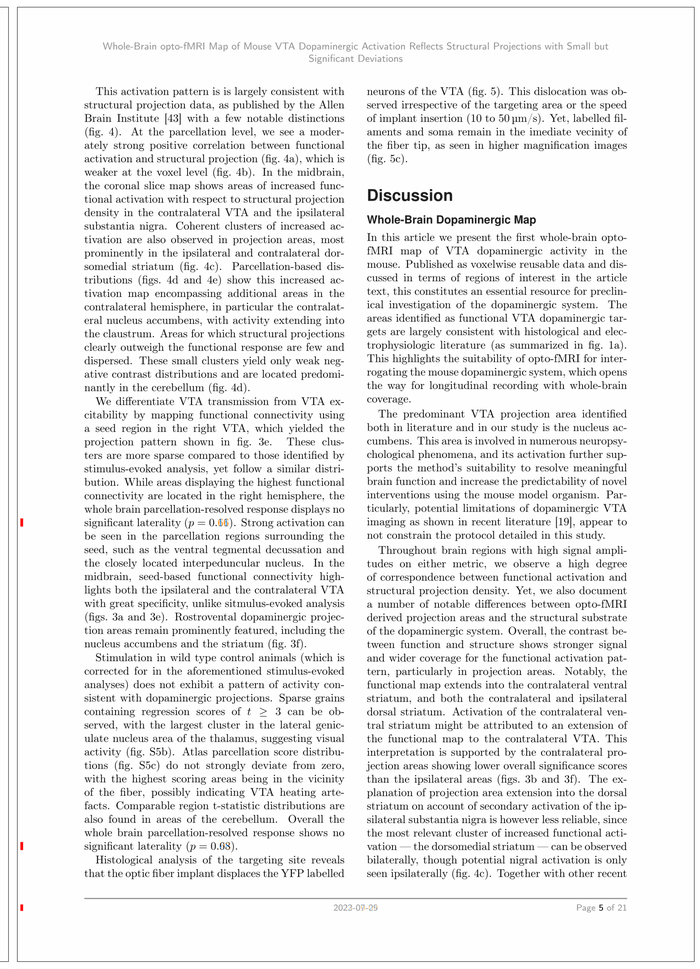
The break-up by phasic and block stimulation is shown in fig. 2 and significance is evaluated accounting for the entire statistical model, consisting of categorical terms for both the stimulus category and the coordinates. The phasic and block levels of the stimulation variable yield p-values of 0.063 and 4.30×10^{-5}, respectively. Upon investigation of the t-statistic map, phasic stimulation further reveals no coherent activation pattern at the whole-brain level (fig. S2b).

The main and interaction effects of the implant coordinate variables are better described categorically than linearly (figs. S1 and 2b). Consequently, the most suitable implant coordinate group for the assay can best be determined on the basis of categorical classification of implant coordinates. We classify the implant coordinates into a "best" and a "rejected" group by k-means clustering the aggregate VTA t-statistic scores into two clusters, and find spatial coherence for the "best" coordinate group (categorization highlighted in fig. 2b).

For block stimulation, the best implant category group (fig. 3a) and the rejected implant category group (fig. 3c) show not only a difference in overall stimulus-evoked signal intensity, but also a difference in efferent distribution, with the rejected implant category efferent spectrum more strongly weighted towards caudal brain areas. This distinction specifically arises for implant categorization based on block scan VTA t-statistic means, and is not as salient if implants are categorized based on a posteroanterior implant coordinate delimiter (fig. S3).

The activation pattern elicited by block stimulation in the best implant category group shows strong coherent clusters of activation. The top activation areas are predominantly located in the right hemisphere, with highly significant laterality ($p = 8.37 \times 10^{-5}$) seen in the comparison of left and right hemisphere atlas parcellation region averages. Activation is seen in regions surrounding the stimulation site, such as the ventral tegmental decussation and the interpeduncular nucleus. The largest activation cluster encompasses well-known dopaminergic VTA projection areas in the subcortical rostroventral regions of the brain (nucleus accumbens, striatum, and the basal forebrain), with weaker activation observed in smaller structures in the vicinity of these regions, such as the fasciculus retroflexus, anterior commissure and the claustrum.

2023-07-25 Page 4 of 21


Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

This activation pattern is is largely consistent with structural projection data, as published by the Allen Brain Institute [43] with a few notable distinctions (fig. 4). At the parcellation level, we see a moderately strong positive correlation between functional activation and structural projection (fig. 4a), which is weaker at the voxel level (fig. 4b). In the midbrain, the coronal slice map shows areas of increased functional activation with respect to structural projection density in the contralateral VTA and the ipsilateral substantia nigra. Coherent clusters of increased activation are also observed in projection areas, most prominently in the ipsilateral and contralateral dorsomedial striatum (fig. 4c). Parcellation-based distributions (figs. 4d and 4e) show this increased activation map encompassing additional areas in the contralateral hemisphere, in particular the contralateral nucleus accumbens, with activity extending into the claustrum. Areas for which structural projections clearly outweigh the functional response are few and dispersed. These small clusters yield only weak negative contrast distributions and are located predominantly in the cerebellum (fig. 4d).

We differentiate VTA transmission from VTA excitability by mapping functional connectivity using a seed region in the right VTA, which yielded the projection pattern shown in fig. 3e. These clusters are more sparse compared to those identified by stimulus-evoked analysis, yet follow a similar distribution. While areas displaying the highest functional connectivity are located in the right hemisphere, the whole brain parcellation-resolved response displays no significant laterality ($p = 0.45$). Strong activation can be seen in the parcellation regions surrounding the seed, such as the ventral tegmental decussation and the closely located interpeduncular nucleus. In the midbrain, seed-based functional connectivity highlights both the ipsilateral and the contralateral VTA with great specificity, unlike stimulus-evoked analysis (figs. 3a and 3e). Rostroventral dopaminergic projection areas remain prominently featured, including the nucleus accumbens and the striatum (fig. 3f).

Stimulation in wild type control animals (which is corrected for in the aforementioned stimulus-evoked analyses) does not exhibit a pattern of activity consistent with dopaminergic projections. Sparse grains containing regression scores of $t \geq 3$ can be observed, with the largest cluster in the lateral geniculate nucleus area of the thalamus, suggesting visual activity (fig. S5b). Atlas parcellation score distributions (fig. S5c) do not strongly deviate from zero, with the highest scoring areas being in the vicinity of the fiber, possibly indicating VTA heating artefacts. Comparable region t-statistic distributions are also found in areas of the cerebellum. Overall the whole brain parcellation-resolved response shows no significant laterality ($p = 0.08$).

Histological analysis of the targeting site reveals that the optic fiber implant displaces the YFP labelled neurons of the VTA (fig. 5). This dislocation was observed irrespective of the targeting area or the speed of implant insertion (10 to 50 µm/s). Yet, labelled filaments and soma remain in the imediate vecinity of the fiber tip, as seen in higher magnification images (fig. 5c).

Discussion

Whole-Brain Dopaminergic Map

In this article we present the first whole-brain opto-fMRI map of VTA dopaminergic activity in the mouse. Published as voxelwise reusable data and discussed in terms of regions of interest in the article text, this constitutes an essential resource for preclinical investigation of the dopaminergic system. The areas identified as functional VTA dopaminergic targets are largely consistent with histological and electrophysiologic literature (as summarized in fig. 1a). This highlights the suitability of opto-fMRI for interrogating the mouse dopaminergic system, which opens the way for longitudinal recording with whole-brain coverage.

The predominant VTA projection area identified both in literature and in our study is the nucleus accumbens. This area is involved in numerous neuropsychological phenomena, and its activation further supports the method's suitability to resolve meaningful brain function and increase the predictability of novel interventions using the mouse model organism. Particularly, potential limitations of dopaminergic VTA imaging as shown in recent literature [19], appear to not constrain the protocol detailed in this study.

Throughout brain regions with high signal amplitudes on either metric, we observe a high degree of correspondence between functional activation and structural projection density. Yet, we also document a number of notable differences between opto-fMRI derived projection areas and the structural substrate of the dopaminergic system. Overall, the contrast between function and structure shows stronger signal and wider coverage for the functional activation pattern, particularly in projection areas. Notably, the functional map extends into the contralateral ventral striatum, and both the contralateral and ipsilateral dorsal striatum. Activation of the contralateral ventral striatum might be attributed to an extension of the functional map to the contralateral VTA. This interpretation is supported by the contralateral projection areas showing lower overall significance scores than the ipsilateral areas (figs. 3b and 3f). The explanation of projection area extension into the dorsal striatum on account of secondary activation of the ipsilateral substantia nigra is however less reliable, since the most relevant cluster of increased functional activation — the dorsomedial striatum — can be observed bilaterally, though potential nigral activation is only seen ipsilaterally (fig. 4c). Together with other recent

2023-07-25 Page 5 of 21

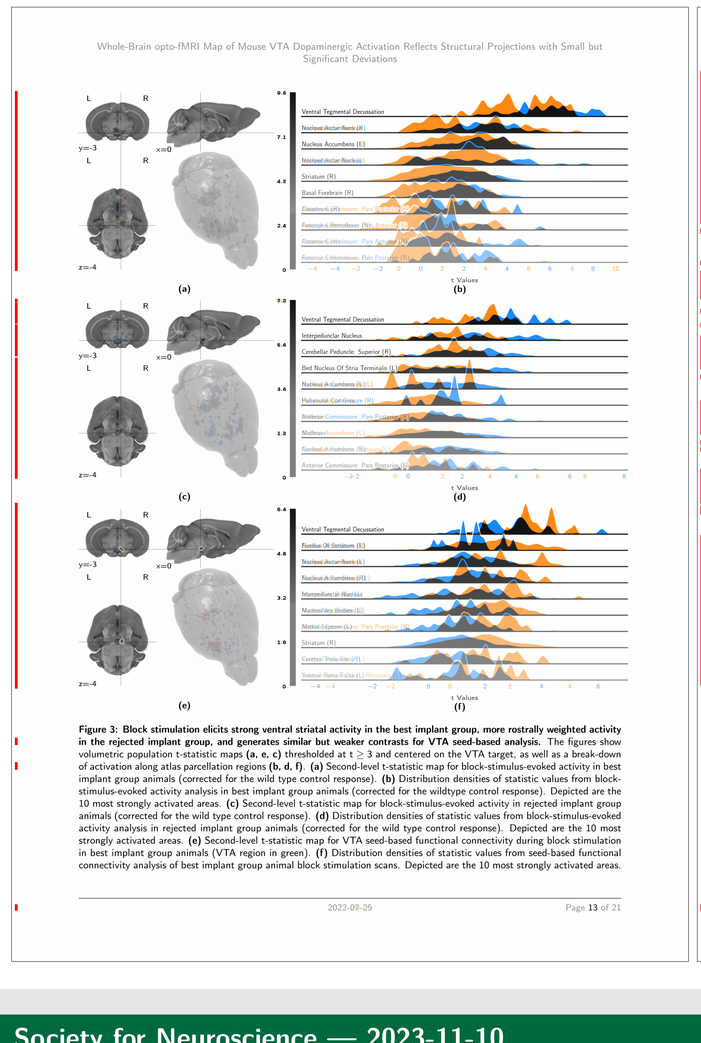

Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

Figure 3: **Block stimulation elicits strong ventral striatal activity in the best implant group, more rostrally weighted activity in the rejected implant group, and generates similar but weaker contrasts for VTA seed-based analysis.** The figures show volumetric population t-statistic maps **(a, e, c)** thresholded at $t \geq 3$ and centered on the VTA target, as well as a break-down of activation along atlas parcellation regions **(b, d, f)**. **(a)** Second-level t-statistic map for block-stimulus-evoked activity in best implant group animals (corrected for the wild type control response). **(b)** Distribution densities of statistic values from block-stimulus-evoked activity analysis in best implant group animals (corrected for the wildtype control response). Depicted are the 10 most strongly activated areas. **(c)** Second-level t-statistic map for block-stimulus-evoked activity in rejected implant group animals (corrected for the wild type control response). **(d)** Distribution densities of statistic values from block-stimulus-evoked activity analysis in rejected implant group animals (corrected for the wild type control response). Depicted are the 10 most strongly activated areas. **(e)** Second-level t-statistic map for VTA seed-based functional connectivity during block stimulation in best implant group animals (VTA region in green). **(f)** Distribution densities of statistic values from seed-based functional connectivity analysis of best implant group animal block stimulation scans. Depicted are the 10 most strongly activated areas.

2023-07-25 Page 13 of 21


Whole-Brain opto-fMRI Map of Mouse VTA Dopaminergic Activation Reflects Structural Projections with Small but Significant Deviations

Figure 4: **Comparing VTA functional activation to structural projection data reveals good correspondence, with deviations involving the dorsomedial striatum and the contralateral ventral striatum.** Depicted are correlation analyses **(a, b)** of the population-level functional and structural statistic scores, alongside statistic distributions **(c, d, e)** for the contrast, taking into account variability across subjects. **(a)** Region-wise regression plot between functional and structural projection maps. Tinted area indicates the 99 % confidence interval of the regression estimate. **(b)** Voxel-wise regression plot between functional and structural projection maps. Tinted area indicates the 99 % confidence interval of the regression estimate. **(c)** Coronal slices, showing the population-level contrast t-statistic between VTA functional activation and VTA structural projections. **(d)** Distribution densities of t-statistics, showing the regions where VTA structural projection exceeds functional activation most strongly. **(e)** Distribution densities of t-statistics, showing the regions where VTA functional activation exceeds structural projection most strongly. **Abbreviations:** Ant. (Anterior), EC (Endopiriform Claustrum), Int. (Intermediate), Med. (Medial), Nc. (Nucleus), p. (Pars), Post. (Posterior), WM (White Matter).

2023-07-25 Page 14 of 21

DOI: 10.5281/zenodo.10085170

uni@chymera.eu