nature neuroscience

TECHNICAL REPORT

https://doi.org/10.1038/s41593-017-0027-7

Corrected: Publisher Correction

An interactive framework for whole-brain maps at cellular resolution

Daniel Fürth[1], Thomas Vaissière[2], Ourania Tzortzi[1], Yang Xuan[1], Antje Märtin[1], Iakovos Lazaridis[1], Giada Spigolon[1], Gilberto Fisone[1], Raju Tomer[3], Karl Deisseroth[3], Marie Carlén[1], Courtney A. Miller[2,4], Gavin Rumbaugh[2] and Konstantinos Meletis[1*]

To deconstruct the architecture and function of brain circuits, it is necessary to generate maps of neuronal connectivity and activity on a whole-brain scale. New methods now enable large-scale mapping of the mouse brain at cellular and subcellular resolution. We developed a framework to automatically annotate, analyze, visualize and easily share whole-brain data at cellular resolution, based on a scale-invariant, interactive mouse brain atlas. This framework enables connectivity and mapping projects in individual laboratories and across imaging platforms, as well as multiplexed quantitative information on the molecular identity of single neurons. As a proof of concept, we generated a comparative connectivity map of five major neuron types in the corticostriatal circuit, as well as an activity-based map to identify hubs mediating the behavioral effects of cocaine. Thus, this computational framework provides the necessary tools to generate brain maps that integrate data from connectivity, neuron identity and function.

The structural and functional mapping of the brain and neural circuits is currently a major endeavor in neuroscience[1,2]. Large projects have been initiated to map the mouse brain in terms of cell types and their activity, long-range connectivity patterns and microcircuit connectivity[3]. Examples of projects to map connectivity and cell types include the Mouse Brain Architecture project[4], the *Allen Mouse Brain Connectivity Atlas*[5] and the Mouse Connectome project[6]. A challenge central to all such large-scale efforts is the need to develop and implement standardized systems to collect, analyze, visualize and share whole-brain data[7]. Advances in experimental methods for the dissection of connectivity and function through for example genetic labeling of connections and neuronal activity at a whole-brain scale now enable the detailed mapping of circuits[8,9]. As a consequence, in addition to the large-scale collaborative projects, an increasing number of individual laboratories have initiated ambitious brain-mapping projects[10–15]. However, the lack of standardized and accessible computational tools often limits the power and feasibility of large-scale whole-brain mapping efforts.

Here we provide an open source software solution to support whole-brain mapping efforts to generate, organize and share whole-brain mapping projects derived from light microscopy. We have developed WholeBrain (http://wholebrainsoftware.org/) to allow investigators to quantify and spatially map multidimensional data from whole-brain experiments and to compare results across experiments in a single standardized anatomical reference atlas. We also developed Openbrainmap (http://openbrainmap.org) to support visualization and sharing of data within and between laboratories in an interactive web-based framework. The whole computational framework is designed to be robust and flexible, allowing its application to a wide variety of imaging systems (for example, widefield, confocal, light-sheet) and labeling approaches (for example, fluorescent proteins, immunohistochemistry and in situ hybridization). Together, this computational framework offers a wide range of tools: powerful image-processing pipelines for mapping labeled neurons in a standardized brain atlas, Bayesian statistical packages handling nested hierarchical data and a framework for producing interactive representations of neuroanatomical data. The framework is available as an open-source R package[16], and we have produced intuitive and concrete guidance through step-by-step video tutorials, allowing for rapid implementation of the system in a standard laboratory environment. Here we demonstrate how WholeBrain and Openbrainmap can be employed to discover brain structure–function relationships by integrating multidimensional anatomical, molecular and functional datasets.

Results

Vector and raster representation in neuroanatomy. All maps are the result of efforts to visually represent the physical world (Fig. 1a). A comprehensive neuroanatomical framework must be able to handle any type of data that can be described as spatially located within the volume of the brain, just as a geographical information system is used to analyze and represent different types of data mapped onto the earth's surface.

In general, maps can be constructed using either raster or vector graphics (Fig. 1b,c). Raster graphics represent data in a grid-cell structure (pixels or voxels) parceled into a row and column matrix (Fig. 1b). This raster-based analysis is the norm in neuroscience today, exemplified by approaches to processing whole-brain calcium imaging[17] statistical parametric mapping in functional MRI[18], differential gene expression, as in the *Allen Gene Expression Atlas*[19], and analysis of whole-brain c-fos expression in behavioral conditions[15] (additional comparisons of recent approaches to mouse brain mapping can be found in Supplementary Table 1). We have aimed to construct a comprehensive framework to represent data at cellular resolution, based on vector graphics, to allow mapping of features through the use of points and curves (Fig. 1d–f). The advantage of this approach is that data can be represented simultaneously in multiple coordinate reference systems—image pixels,

[1]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. [2]Department of Neuroscience, Scripps Research Institute, Jupiter, FL, USA. [3]Department of Bioengineering, Stanford University, Stanford, CA, USA. [4]Department of Molecular Medicine, Scripps Research Institute, Jupiter, FL, USA. *e-mail: dinos.meletis@ki.se; furth@cshl.edu

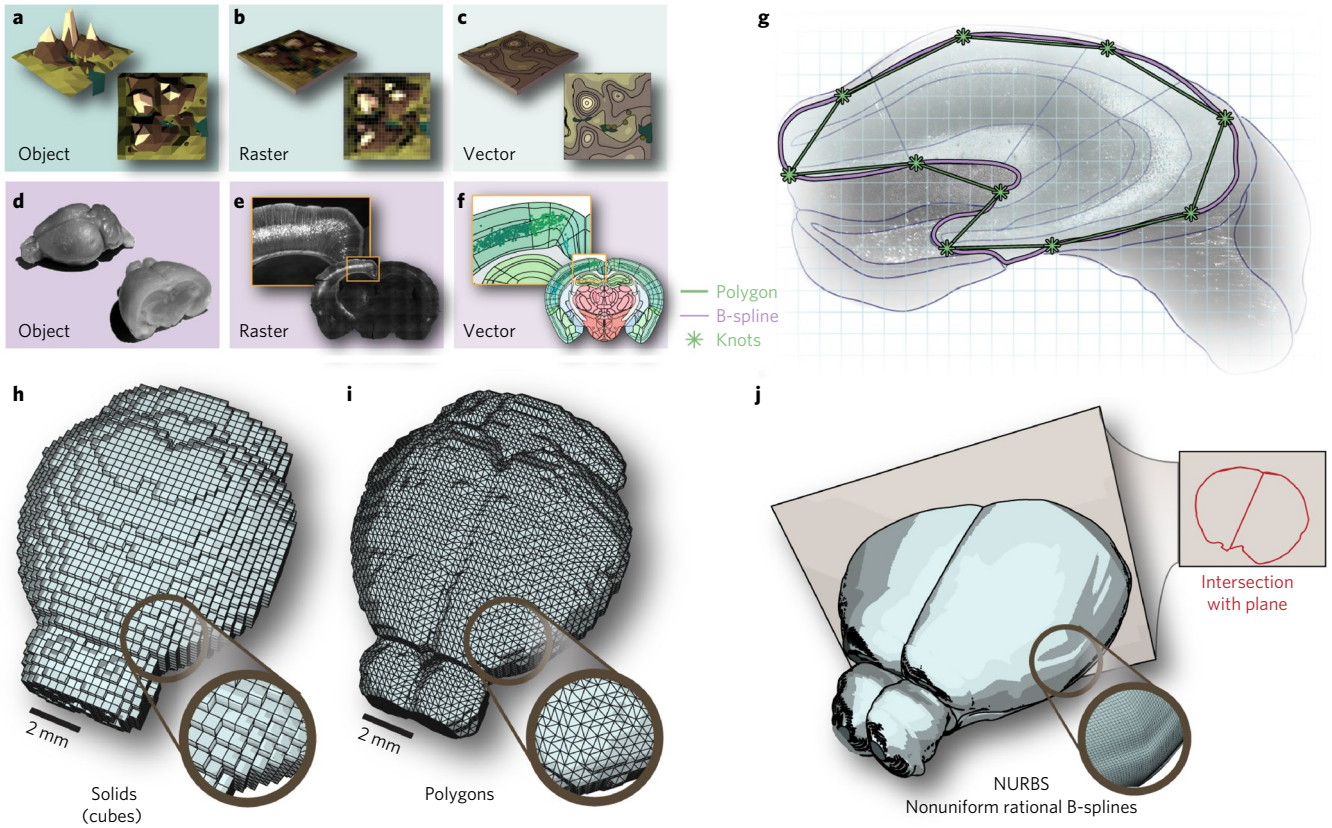

Fig. 1 | A reference atlas based on vector graphics. a–f, In cartography, an object (**a**) is usually represented either in raster format (**b**) or vector format (**c**). Similarly, brain tissue (**d**) can be represented as raster images (**e**) or as an image composed of geometric points and curves in a vector format (**f**). **g**, Brain regions defined by polygons (green) are nonscalable compared to regions obtained from nonuniform rational B-splines (NURBS, purple). **h–j**, In 3D, reference atlases can be defined based on different geometrical primitives: (**h**) voxels can be reduced to primitive solids such as cubes, (**i**) polygon surfaces provide a more compact representation of the surface only, and (**j**) NURBS provide a representation of the surface based on smooth B-splines where intersections in the form of curves can be computed to an arbitrary cutting plane.

stereotactic coordinates and actual tissue dimensions—instead of only transforming data into a single spatial reference frame.

Scale-invariant reference atlas. Tissue coordinates of individual features (for example, neurons) are normally mapped into a standardized reference atlas using raster or polygon graphics[20]. We have instead generated a reference atlas of the adult mouse brain based on smooth curves, which can enhance the possibilities for efficiently representing, sharing and editing anatomical definitions in a standardized reference atlas.

We therefore decided to represent brain regions as smooth surfaces using nonuniform rational B-splines (Fig. 1g). To generate a three-dimensional (3D) anatomical reference atlas, we lofted points between adjacent coronal sections. Compared to existing reference atlases based on solids (Fig. 1h) or polygons (Fig. 1i), this approach offers a number of advantages. First, the atlas is scale-invariant and can be easily applied on images of any pixel resolution. This feature enables comparison of image data across microscope systems. Second, atlas intersections can be computed at arbitrary angles (Fig. 1j). This feature supports projects in which visualization of regions and pathways is better achieved at angles other than the canonical (for example, coronal or sagittal). Third, the topological representation of brain regions is data-efficient and project results can be efficiently shared in a small format. Fourth, the nonuniform rational B-splines format enables editing of the brain atlas to accommodate customization and atlas improvements.

The reference atlas is based on the neuroanatomical definitions found in the Allen Institute mouse reference atlas, thereby integrating the available neuroanatomical and molecular data[5,20]. As an example, this feature enables queries of gene expression from the Allen in situ hybridization atlas for any given stereotactic coordinate. In addition, the reference atlas can easily interface with other applications, for example, by standardization of tissue sectioning using custom 3D-printed brain blocks (Supplementary Fig. 1a–d) or through generation of polygon-shaped areas for automated laser capture microdissection of defined brain nuclei (Supplementary Fig. 1e–j).

Interactive web interface. We further developed an interactive web interface to allow investigators to rapidly and in a standardized format visualize mapping results and to allow them to share data with collaborators. The aim was to develop an interface in which data visualization could be performed directly in a standard Internet browser and without setting up a server. In this interface, detected features such as cell bodies and the fit of the reference atlas to the tissue can be visualized as an overlaid layer of vector graphics on the original raster image (Fig. 2a–c; see http://openbrainmap.org/example_section/ for an example), queried directly as tabulated summary statistics (Fig. 2d,e) or rendered in stereotactic coordinates within the reference atlas (Fig. 2f).

To enable rapid visualization and data sharing, output is provided as a browser-based HTML file with associated image tiles. As an example, raster data from a single raw image of a coronal

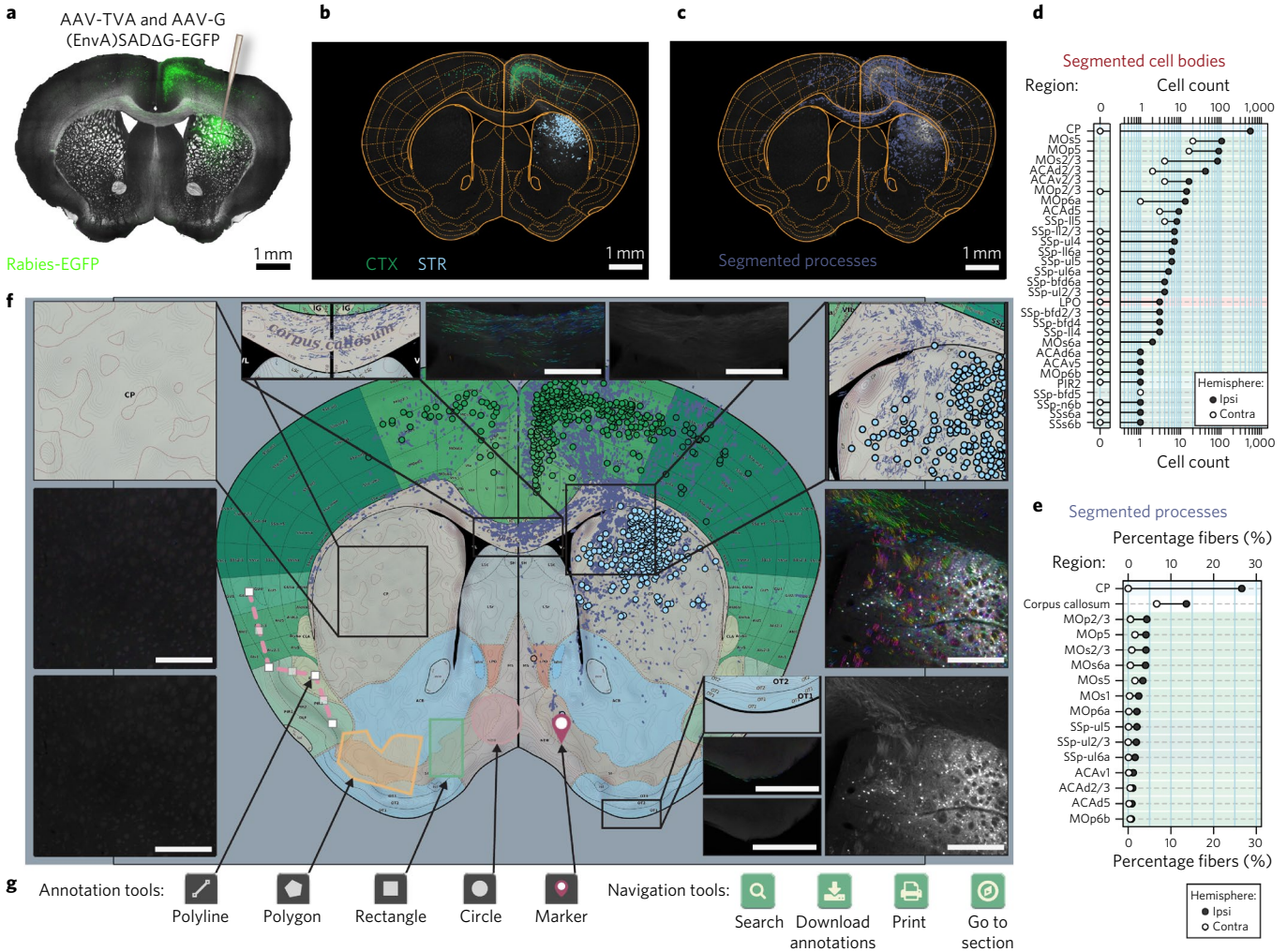

Fig. 2 | A framework for standardizing and sharing neuroanatomical data. a, Coronal section from D1-Cre mouse with rabies tracing targeted to the dorsal striatum (right hemisphere: green, rabies-EGFP; gray, bright-field). **b**, Registration of atlas regions as well as segmentation of 1,108 individual cell bodies, color-coded according to region. Striatum, STR; cortex, CTX. **c**, Segmented rabies-EGFP$^+$ processes. **d**, Quantification of segmented cell bodies. Contra, contralateral; ipsi, ipsilateral. **e**, Quantification of segmented processes. **f**, Web-based vector representation of results registered to reference atlas. Insets: close ups of segmentation results transformed into reference atlas (top or left), compared to segmentation results of fiber tracts with orientation color-coded by hue (middle) as well as original fluorescent image (bottom or right). Scale bars, 500 µm. **g**, Interactive web-based annotation and navigation tools. Brain regions abbreviated in **d** and **e** are defined in Supplementary Fig. 6.

brain section (16 bit, 300–400 Mb) is represented as a set of JPEG tiles of different resolutions (approximately 20 Mb in total). For the purpose of transmitting segmentation and registration results over asynchronous browser–server communication, we constructed an open standard format for spatial features, along with their nonspatial attributes, based on JavaScript Object Notation (see Methods). The web interface is designed with mobile-first principles, supporting use on mobile devices, and offers users the ability to further draw, edit or define regions of interest (Fig. 2g).

Segmentation by multiresolution decomposition. Segmentation is the process of assigning labels to distinct parts of an image to mark features of interest, such as labeled cell bodies and fiber tracts. In large mapping projects, it is essential to implement automated approaches to feature segmentation that can be applied to all images with minimal manual intervention.

We used multiresolution decomposition[21] as an approach to automatically segment fluorescently labeled cell bodies and fiber tracts. In this segmentation approach, features of different size (for example, cell bodies and cell processes) can be detected by their mapping on a distinct spatial scale. As an example, an apical dendrite (~1–2 µm thick) is defined on a smaller spatial scale than a cell body (~10–20 µm in diameter). Genetic labeling of dendrites and cell bodies can therefore be decomposed by designating segmentation of these features on separate scales.

To decompose the image into different spatial scales (Fig. 3a,b), the image data is sent through a stationary wavelet filter bank composed of a set of high-pass and low-pass filters (Supplementary Fig. 2a,b). At each iteration, the output of the high-pass branch is stored as a set of coefficients called detail coefficients (Fig. 3c–e). Detail coefficients contain information on the correlation between a wave-like function (wavelet) at a specific scale and the fluorescent signal at a specific location in the image. Additional decompositions at lower resolution levels are obtained by iterating the procedure on the output of the low-pass branch. The output data of the low-pass branch at each iteration are called approximation coefficients because, at each level, the approximation to the image at that resolution level becomes coarser; the detail coefficients

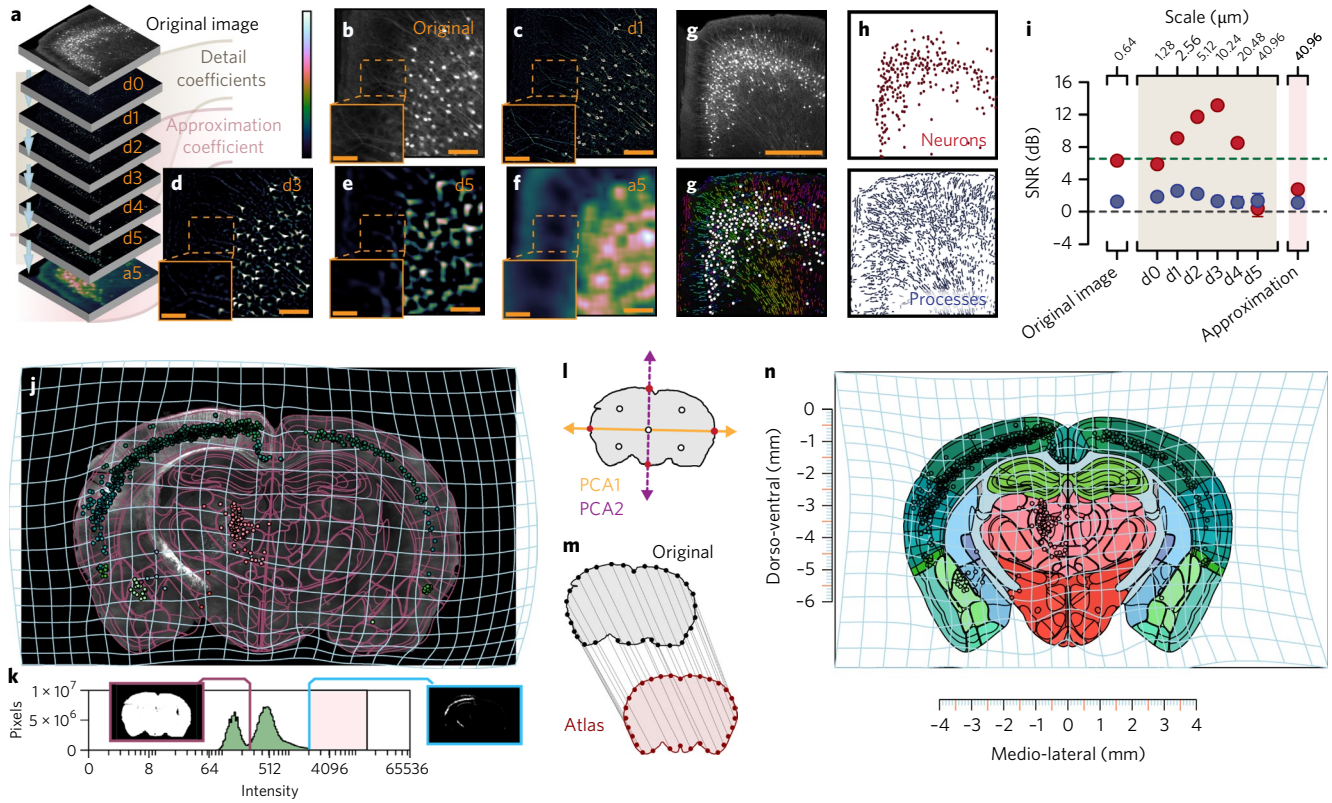

Fig. 3 | Method for segmentation and registration. a, Wavelet multiresolution decomposition. **b**, Original image tile. Scale bars, 100 μm. **c**, Detail coefficients *d1* at a scale period of 1.28 μm. **d**, Coefficients at *d3* with scale period 10.24 μm. **e**, Coefficients at *d5* with scale period 40.96 μm. **f**, Approximation coefficients. Scale bars, 100 μm. Inset scale bars, 50 μm. **g**, Top: original 16-bit image tile. Bottom: segmented processes with their direction color-coded by hue, as well as cell bodies coded as white circles. Scale bar, 500 μm. **h**, Top: segmentation result separated into cell bodies. Bottom: segmented process. **i**, Comparison of binary segmentation on different detail coefficients to a human annotator ($n = 273$ cell bodies; red, cell bodies; blue, processes; gray dashed line, SNR of 1:1; green dashed line, Rose criteria SNR 5:1[33]). Error bars: 95% confidence intervals around the mean. **j**, Image section of rabies-EGFP targeted to dorsal striatum in D2-Cre mouse (purple lines, transformed reference atlas; light blue grid, backward warp transform from atlas to original image). **k**, Segmentation of brains section based on autofluorescence (vertical purple line). Pink shaded area indicates range of fluorescent intensity where cell bodies can be found. **l**, Correspondence generation by principal components (red dots, intersection with contour; white dots, midpoints). **m**, Thirty-two correspondence points between original contour (gray) and reference atlas (red). **n**, Forward warp transform (light blue grid) of segmentation output into stereotactic space.

indicate what detail in the image is lost (Fig. 3a–f). We extracted individual features by thresholding the fluorescence intensity for the detail coefficients of interest (i.e., *d3* for cell bodies; Fig. 3c–e). We then extracted each feature by applying a standard connected-components algorithm[22], which groups connected pixels by their borders (Fig. 3g,h).

To address the sensitivity of the approach, we evaluated the signal-to-noise ratio (SNR) in manually defined regions of interest (neurons versus processes) after applying multiresolution decomposition and compared this to binary thresholding of the fluorescent signal in the original image (Fig. 3g). We found that the SNR for segmentation of neurons reached a peak at a theoretical sampling scale period of 10.24 μm, approximately the size of a cell body (Fig. 3i). Similarly, we found that the SNR for processes reached a peak at a sampling period of 2.56 μm (i.e., the approximate diameter of an apical dendrite). In summary, detail coefficients at separate spatial scales can be used to identify genetically marked cell bodies and processes.

In addition, multiresolution decomposition is particularly valuable for comparison and integration of imaging results obtained from different microscope systems. In essence, features segmented from low- and high-resolution images can be directly compared by analyzing detail coefficients at scales with similar scale periods (Supplementary Fig. 2c–i). Multiresolution decomposition therefore supports data gathering and comparison of segmentation results across imaging platforms, projects and laboratories.

Registration to standardized atlas. Image registration is the process by which sets of images (for example, coronal mouse brain sections) are transformed into another coordinate system (for example, a stereotactic coordinate system or a reference atlas). Traditional registration algorithms in neuroimaging follow a voxel-based analysis[15,19,23]. Registrations using voxel-based atlases, such as the *Allen Mouse Brain Reference Atlas*[20] or Waxholm space[24], are limited to 21.5–25 μm, and even newer isotropic renditions such as the next-generation common coordinate framework (CCF, v3)[25] are still limited to a voxel resolution of approximately 10 μm. To decrease computational time, these registration approaches most often require matching of image resolutions by upsampling the lower-resolution atlas and downsampling the imaged brain sections. In contrast, we decided to perform image registration by segmenting out the contour of the brain section using the inherent autofluorescence of the brain section itself (Fig. 3j,k). In this way, image registration does not introduce artifacts derived from pixel interpolation associated with upsampling the reference atlas. In addition, it is not necessary to dedicate one imaging channel for the registration process, which reduces image acquisition time and facilitates fluorescent multichannel experiments (multiplexing).

To generate landmarks between the brain tissue and the reference atlas, the segmented contour of the imaged brain section is reduced into a set of points[26]. To achieve this, the two initial principal components are first extracted from the positions of the

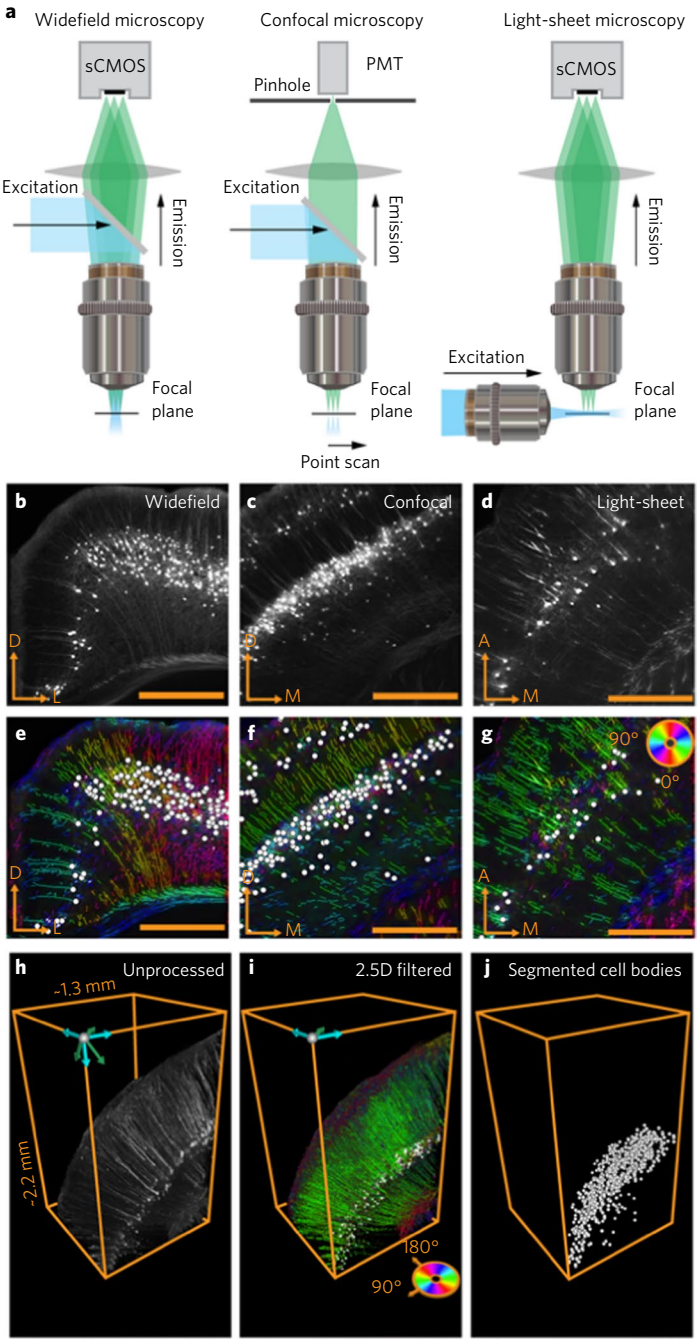

Fig. 4 | Compatibility of the framework with different imaging systems. a, Compatibility of the framework with different imaging systems, including widefield, confocal and light-sheet fluorescent microscopy. sCMOS, scientific complementary metal-oxide semiconductor camera; PMT, photomultiplier tube. **b**–**d**, Raw fluorescent images of rabies-EGFP for (**b**) widefield, (**c**) confocal and (**d**) light-sheet microscopes. **e**–**g**, Segmentation result of both cell bodies (white circles) and processes (color coded according to angle) for (**e**) widefield, (**f**) confocal and (**g**) light-sheet images in **b**,**c**. D, dorsal; L, lateral; M, medial; A, anterior. **h**, Cortical neurons from a volume imaged with light-sheet microscopy in a Thy1-eYFP mouse. **i**, Segmentation algorithm applied across *z*-stacks (2.5D filter). **j**, Cell bodies individually segmented from fluorescent signal attributed to processes and fiber tracts.

pixels defining the tissue section contour (Fig. 3l). Next, intersection points between the principal components and the tissue contour are extracted and stored as the initial first level of reference points (Fig. 3m). By default, four iterations ($n=4$) are executed, resulting in $32=2^{n+1}$ reference points along the contour of the brain section (Supplementary Fig. 3). The two sets of reference points (atlas and tissue-section reference points) constitute a set of correspondence points that align the atlas and the tissue section (Fig. 3m). These correspondence points are then pruned[27] to minimize effects of tissue damage or distortions on the registration result. These correspondence points are used as input to solve analytically the minimization of the bending energy of a thin-plate splines deformation field[28]. The deformation field can then be used to generate a mapping between every coordinate in the brain tissue section (Fig. 3j) to that of the reference brain atlas (Fig. 3n). Using this approach, a collection of individual brain sections can be reconstructed into a 3D representation of the imaged data (Supplementary Video 1).

When the automated registration results in incorrect placement of region boundaries, it is possible to change or remove correspondence points manually and rerun the thin-plate splines algorithm (Supplementary Fig. 4). To estimate how the measurement accuracy depended on the angle of tissue sectioning and the registration plane, we compared detection of labeled neurons obtained in a number of regions from brains sectioned on the coronal or sagittal planes (Supplementary Fig. 5). As a measurement of registration accuracy, we defined the difference in the percentage of labeled cells found in each anatomical subregion in sagittal compared to coronal reconstructions. The average error across all anatomical regions was found to be 0.29% (standard deviation (s.d.) = 0.96%, $n = 239$ brain regions), as defined by the root-mean-squared error (Supplementary Figs. 6 and 7). Notably, reconstructed 3D brains can be resliced at arbitrary angles, although the sampling strategy will still impose limitations on the spatial resolution that can be obtained (Supplementary Fig. 8).

Fiber-tract tracing. To construct mesoscale connectomes, it is necessary to perform fiber-tract tracing[5]. Fiber tracts are geometrical features extended in space (spatially coherent) with a set of orientations at different locations. We used the properties of coherency and orientation to further increase the SNR for segmentation of fiber tracts and cell processes by computing the structure tensor on the detail coefficients of interest followed by segmentation of topological skeletons (see Methods). The segmented topological skeletons, together with the associated orientation of each pixel (Fig. 4e–g), can be used as input to tertiary tractography approaches[29].

Resolving the identities of fiber tracts across sectioned tissue is challenging. It is therefore desirable to perform fiber tracing in 3D, for example, in intact clarified brain tissue imaged using confocal or light-sheet microscopy (Fig. 4a–j). To process 3D image stacks (Fig. 4h), we developed a set of 2.5D algorithms that segment out processes and cell bodies along the *z*-stack plane (Fig. 4i). This approach is comparable in speed to 2D image processing but provides the accuracy of running 3D wavelet filters by running an additional connected-components algorithm across the *z*-stack after combined wavelet and structure tensor on each *z*-plane.

Mapping of single cell co-expression. We further applied our approach to determine, at the single-neuron level, the expression of markers after imaging of fluorescent proteins (for example, EGFP) and immunohistochemistry. To map the anatomical distribution of neuron types, we chose to determine the spatial distribution of interneurons using transgenic labeling in combination with immunohistochemistry. We used a transgenic mouse line (Lhx6::EGFP) in combination with detection of parvalbumin-expressing (Pvalb$^+$) and neuropeptide Y-expressing (NPY$^+$) interneurons. Automated image analysis resulted in the segmentation of 9,021 labeled cells in a single coronal mouse brain section (Fig. 5): Lhx6::EGFP: 66% ($n = 5{,}958$), PVALB::Cy5: 63% ($n = 5{,}701$) and NPY::Cy3: 15% ($n = 1{,}342$). The processing time for segmentation and registration to the atlas was approximately 30 s, indicating the scalability of whole-brain mapping projects based on our framework. Further demonstrating the advantages of an automated framework, we could, with high resolution, assign the position of each of these 9,021 labeled cell bodies to 116 unique regions (Fig. 5a–d). This type of comprehensive single-neuron mapping generates detailed anatomical definitions that can, for example, describe the relative distribution of neuron types across cortical layers and in different anatomical subregions (Fig. 5e). We confirmed the non-overlapping distributions of PVALB and NPY ($n = 0$) in Lhx6$^+$ cells (Fig. 5g,h; 37% of cortical NPY$^+$ cells co-express Lhx6::EGFP, $n = 238$, and 73% of cortical PVALB$^+$ cells co-express Lhx6::EGFP, $n = 2{,}752$). In contrast to cortex, in the striatum 84% of NPY$^+$ cells were also Lhx6::EGFP$^+$ ($n = 101$).

We further investigated the laminar distribution of the five different neuron clusters and found that PVALB$^+$ (PVALB only) as well as NPY$^+$ (NPY only) cells were primarily localized to layer II/III of primary somatosensory cortex barrel field, whereas Lhx6$^+$ (Lhx6 only) and Lhx6$^+$NPY$^+$ cells were enriched in layer V (Fig. 5e).

Notably, our framework further enables quantification of the average fluorescence intensity of each single cell for all labeling channels, resulting in a quantitative description of expression levels at the single cell level and therefore also of population clustering, in a manner similar to gating in flow cytometry (Fig. 5f). Using the quantitative information for each fluorescent channel, we clustered neurons into five distinct populations based on expected maximization with Gaussian mixture models (Fig. 5h,i). These clusters were then superimposed on the reference atlas to align molecular and anatomical definitions of the map (Fig. 5h). For example, we found that the majority of PVALB$^+$ and Lhx6::EGFP$^-$ neurons (PVALB$^+$Lhx6$^-$) were in the thalamic reticular nuclei (Fig. 5g,i) and that Lhx6$^+$ (Lhx6::EGFP$^+$PVALB$^-$NPY$^-$) cells were, for example, found preferentially in zona incerta and in medial amygdalar nucleus. This approach can similarly quantify gene expression on the subcellular level through detection of single mRNA molecules with fluorescent in situ hybridization (Supplementary Fig. 9).

Tracing connectivity using modified rabies virus and transgenic mice. Whole-brain mapping of monosynaptic connectivity using modified rabies viruses can generate large datasets of cell-type-specific connectomes at the mesoscale level[18,30,31]. We applied rabies virus tracing to identify the presynaptic partners of defined neuronal subtypes in the corticostriatal pathway (Supplementary Video 2). We used a Cre-mediated genetic targeting approach to limit the uptake of rabies virus to defined neuron types in Cre-expressing mouse lines. We first mapped the whole-brain inputs to excitatory or inhibitory neurons in the motor cortex (MOp) using Camk2a-Cre or Gad2-Cre mice, respectively (Fig. 6a–d). To target the main neuron types in striatum (CPu), we used D1-Cre, D2-Cre and Chat-Cre mice to target either medium spiny neurons of the direct pathway or indirect pathway and cholinergic interneurons, respectively (Fig. 6e–j). We mapped the precise neuroanatomical distribution of the labeled presynaptic neurons at a whole-brain scale, which produced a rich dataset that included whole-brain tracing data from 15 mice and 349,959 annotated neurons (Fig. 6k,l).

Quantitative comparison across a large number of animals and experiments allowed us to search for cortical regions that showed connectivity patterns similar to those of the thalamus (i.e., innervating both MOp and CPu), which could then be anatomically and functionally described as contextual units[32,33] instead of as input units to motor cortex. We therefore analyzed the whole-brain tracing dataset to specifically identify cortical subregions that exhibited dense targeting of both MOp and CPu. We identified three candidate regions that satisfied the criteria of serving as contextual units: primary somatosensory cortex (SSp), anterior cingulate area and orbital cortex (ORB). We further analyzed the cell-type-specific organization of SSp, anterior cingulate area and ORB inputs to striatum and found that the ipsilateral inputs in ORB (ORBvl and ORBl) displayed preferential targeting to D1$^+$ striatal cells (Fig. 6m). ORB therefore displays a unique connectivity pattern among the corticostriatal network, preferentially targeting striatal D1$^+$ neurons.

To investigate the laminar distribution of the corticostriatal organization further, we explored our dataset to identify proposed cell-type-specific connectivity patterns. Unlike previous studies that have relied on a small number of cortical neurons[34,35], we quantified a large number of presynaptic neurons in the corticostriatal pathway (320,055 neurons) after monosynaptic rabies tracing from distinct striatal subtypes. We analyzed the laminar distribution of cell-type-specific corticostriatal inputs in SSp and found that the major inputs to D1$^+$ and D2$^+$ striatal neurons were found in layer

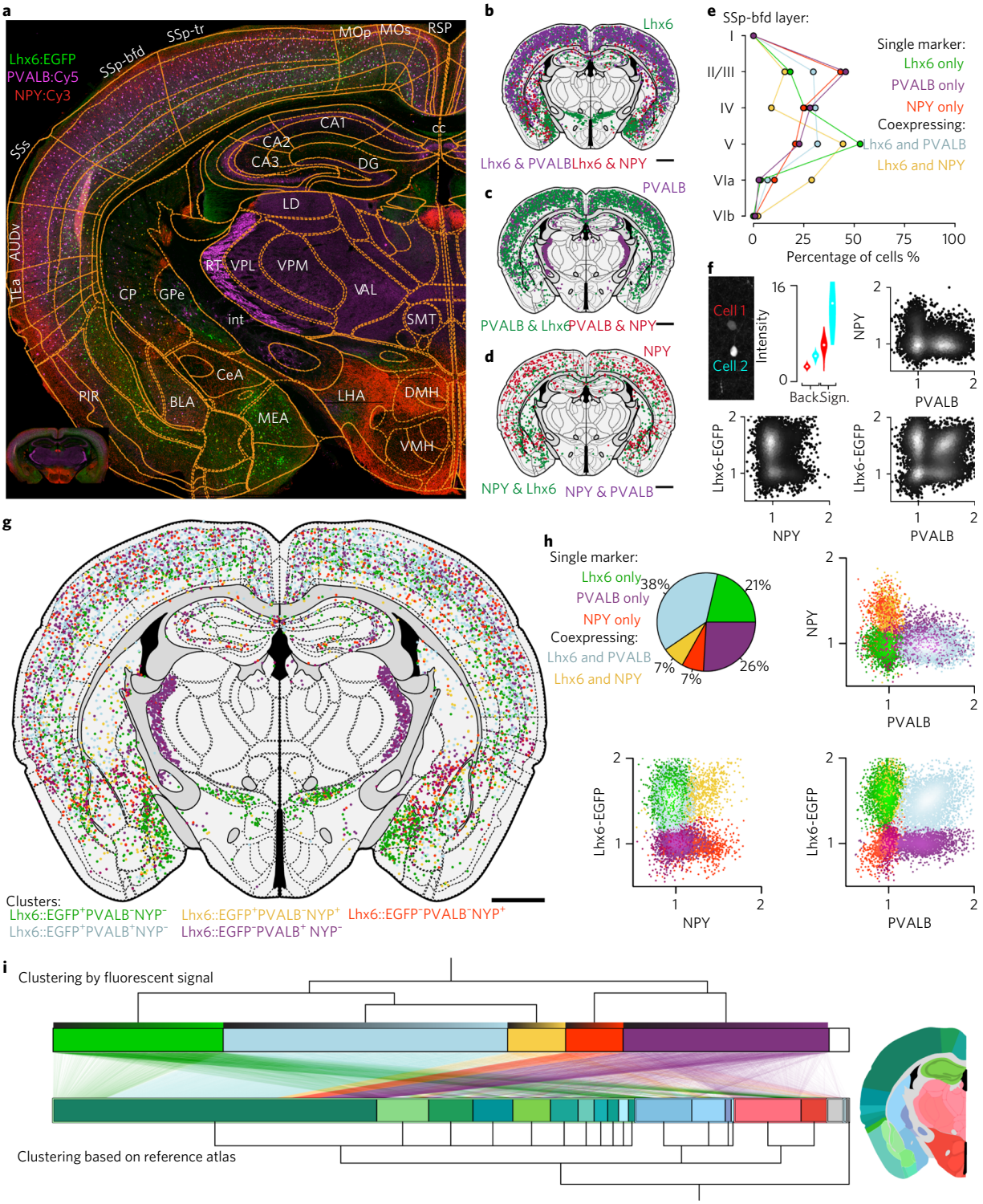

Fig. 5 | Mapping of molecular identity and neuron types. a, Section of an Lhx6::EGFP transgenic mouse (green) stained for parvalbumin (magenta, PVALB::Cy5) and neuropeptide Y (NPY::Cy3). Abbreviations provided in Supplementary Fig. 7. **b**, Segmented Lhx6::EGFP$^+$ cells and their co-expression with either PVALB (purple) or NPY (red); green denotes Lhx6::EGFP$^+$ neurons negative for both NPY and PVALB. **c**, PVALB::Cy5$^+$ cells and their co-expression with either Lhx6::EGFP (green) or NPY (red); no co-expression was found between PVALB and NPY. **d**, NPY::Cy3$^+$ cells (red) and their co-expression with Lhx6::EGFP (green). **e**, Laminar distribution of cell types in primary somatosensory cortex barrel field (SSp-bfd). **f**, Definition of molecular signal at the single cell level (background (back.), $n = 400$ pixels; signal (sign.), $n = 16$ pixels). **g**, Registration of 9,021 segmented cell bodies to 116 unique regions in the reference atlas and clustered into five discrete classes based on fluorescent intensity. **h**, Clustering of cells based on fluorescent intensity alone into five discrete populations (green, Lhx6-EGFP$^+$ only; magenta, PVALB$^+$ only; red, NPY$^+$ only; yellow, NPY$^+$Lhx6-EGFP$^+$; light blue, PVALB$^+$Lhx6-EGFP$^+$). Pie chart illustrates percentage overlap. **i**, Clustering based on marker identity (top) compared to clustering based on neuroanatomical location for each cell (bottom). Lines represent individual neurons mapping between the two types of clustering (molecular identity versus anatomical). Neurons are sorted in ascending order according to their relative fluorescence intensity in each cluster (top gradient on each cluster). Anatomical color-coding is shown in the representative section, –1.5 mm from bregma (scale bar, 1 mm).

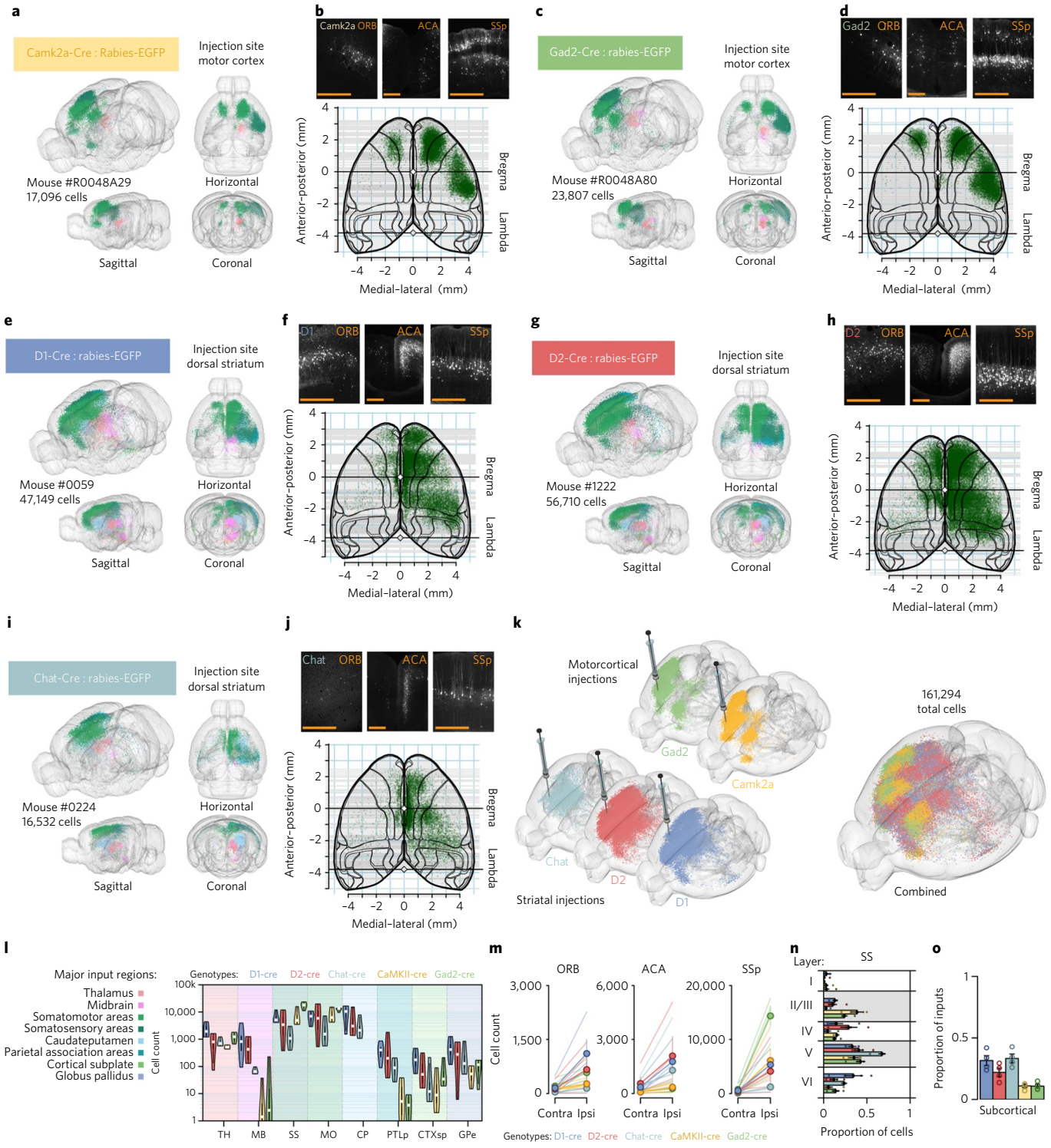

Fig. 6 | Retrograde monosynaptic tracing of corticostriatal networks. a, Reconstruction of inputs to Camk2a neurons in motor cortex by targeted injection of glycoprotein-deleted EGFP$^+$ EnvA pseudotyped rabies virus, SADΔG-EGFP(EnvA). EGFP$^+$ neurons are color-coded based on anatomical location; see **l**. **b**, Top: cortical input to Camk2a neurons in motor cortex in ORB, anterior cingulate area (ACA) and SSp. Bottom: cortical overview of inputs from cortical cells roughly above −2.25 mm dorsoventral from the midline cortical surface. **c**,**d**, As in **a** and **b** but for Gad2 neurons in motor cortex. **e**,**f**, As in **a** and **b** but for D1$^+$ medium spiny neurons in dorsal striatum. **g**,**h**, As in **a** and **b** but for D2$^+$ medium spiny neurons in dorsal striatum. **i**,**j**, As in **a** and **b** but for cholinergic interneurons in dorsal striatum. **k**, Illustration of rabies-EGFP-labeling in individual brains (from five different transgenic Cre lines) used for cell-type-specific input comparisons. The large brain is a composite in which all neurons are combined (n = 161,294 neurons across 5 mice). **l**, Violin plot of monosynaptic inputs from major input regions (n = 4 mice for each genotype). White circles show the medians; thick black lines indicate the first and third quartiles; thin black lines extend 1.5× the interquartile range; polygons represent density estimates of data and extend to extreme values. **m**, Laterality of cortical inputs in specific regions (ORB, ACA, SSp). **n**, Layer-specificity of monosynaptic inputs from SSp. Colors as in **l**. **o**, Proportion of inputs from subcortical regions. Circles show individual mice. Colors as in **l**. Error bars: ±1 standard error of measurement. Scale bars, 500 µm.

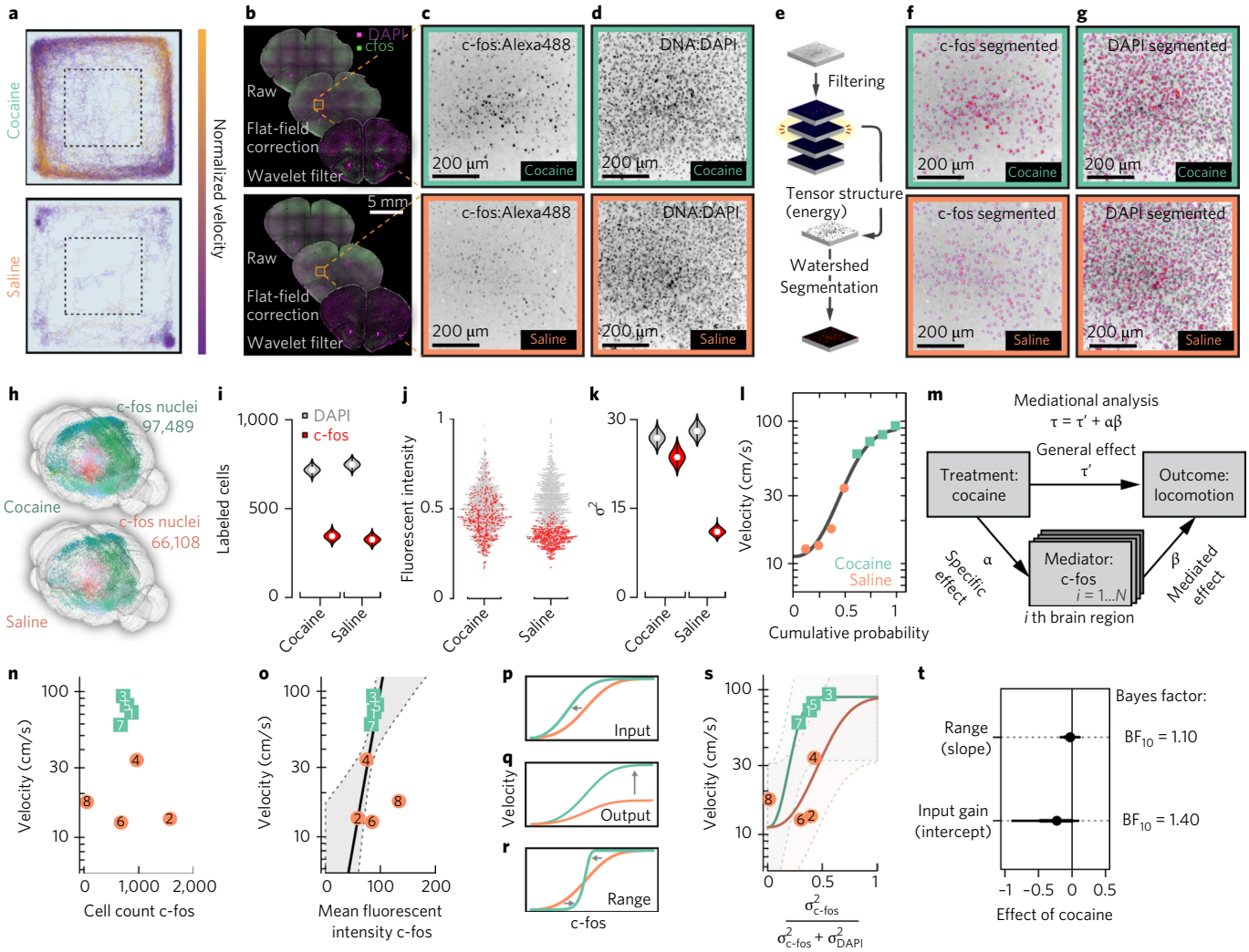

Fig. 7 | Decoding motor behavior by immediate early gene activity. a, Behavioral track tracing from the open field test in two mice injected with either cocaine or saline. **b**, Coronal sections from the same mice showing the major steps of preprocessing before segmentation. **c**, Close up on ORB and c-fos Alexa Fluor-488 (Alexa488) staining. **d**, As in **c** but for DAPI. **e**, Postfiltering segmentation steps include computing the tensor structure and performing watershed-based segmentation. **f**,**g**, Segmentation result (red) overlaid on images of (**f**) c-fos and (**g**) DAPI. **h**, Whole-brain 3D reconstruction of same mice as in **a**–**j**. **i**, Violin plot of cell count estimates obtained from ORB as shown in **f** and **g**; white circle marks median; thick black lines indicate the first and third quartiles; thin black lines extend 1.5× the interquartile range; polygons represent density estimates of data and extend to extreme values. **j**, Normalized fluorescent intensity for 2,135 individual nuclei in ORB. **k**, Violin plot of variance (σ^2) estimates of fluorescent intensity in ORB; white circles mark medians; thick black lines indicate the first and third quartiles; thin black lines extend 1.5× the interquartile range; polygons represent density estimates of data and extend to extreme values. **l**, Cumulative distribution of locomotor velocity in open field across 8 mice (gray line, best fit four-parameter Weibull distribution). **m**, Meditational regression analysis for c-fos whole-brain data. **n**, Velocity in open field test as a function of cell count of c-Fos$^+$ nuclei of the orbital cortex in saline-treated (coral) and cocaine-treated (turquoise) mice; numbers indicate animal identification number. **o**, Velocity as function of mean fluorescent c-fos intensity; thick gray line indicates regression line when autofluorescence is added as a covariate (gray area, 80% credible interval). **p**–**r**, Gain control by (**p**) input gain, (**q**) output gain or (**r**) range compression. **s**, Velocity as a function of DAPI-standardized c-fos variance. Lines show posterior estimates of Weibull psychometric functions; dotted lines show 80% credible intervals. **t**, Posterior estimates of the effect of cocaine on slope and intercept. Error bars, 70% (thick) and 80% (thin) credible interval; $n = 4$ mice per group (cocaine, saline).

V, although we also found layer II/III and IV inputs (Fig. 6f,h,n). Inputs to Chat$^+$ striatal neurons were primarily located in layer V and VI in SSp (Fig. 6j,n). In comparison, we found that inputs to Camk2a$^+$ and Gad2$^+$ cells in MOp primarily originated from layer II/III of SSp but could be also found in layer V (Fig. 6n).

Comparing inputs to the striatal neuron subtypes on the whole-brain scale, we found that D2$^+$ striatal neurons display preferential input from neurons in cortical regions compared to inputs from subcortical structures, whereas D1$^+$ and Chat$^+$ striatal cells receive more balanced cortical–subcortical input (Fig. 6o). This evidence supports a discrete cell-type-specific pattern of connectivity both in the corticostriatal pathway as well as in the input balance from cortical versus subcortical regions.

Decoding motor behavior from whole-brain intermediate early gene activity. To investigate the behavioral relevance of the corticostriatal organization identified by monosynaptic rabies tracing, we performed a whole-brain functional assay. Labeling neurons based on their expression of immediate early genes (for example, c-fos) as a proxy measure of neural activity has been valuable in identifying regions and neuron types that are recruited during behavior[36].

Studies have mapped cocaine-induced c-fos expression in cortical and basal ganglia circuits[37,38], and more recently, similar approaches have mapped c-fos at a whole-brain scale[13,14]. We selected acute cocaine administration as an assay to map how motor function recruits distinct corticostriatal circuits. We exposed mice to cocaine or saline and then imaged c-fos expression to identify neurons and regions that displayed activation correlated with increased motor activity. Motor responses of mice injected with 20 mg/kg cocaine or saline were monitored for 20 min in an open field (Fig. 7a; $n=4$ mice per group), and brains were sectioned ($M=93.6$ sections per brain, s.d. = 11.2 sections per brain) and stained for c-fos as well as DAPI (Fig. 7b–d).

To segment cell nuclei and determine their c-fos fluorescence intensity, we computed the tensor energy on detail coefficients for subsequent watershed segmentation (Fig. 7e–g and Supplementary Fig. 9). The segmentation algorithm is based on spatial frequency analysis, allowing segmentation of nuclei independently of their signal intensity, which resulted in detection of c-fos signal over a wide dynamic range (Fig. 7c–f). We mapped the position of c-fos$^+$ neurons at the whole-brain scale (Fig. 7h).

We focused our analysis on activity mapping to ORB based on the corticostriatal connectivity pattern. We found that cocaine did not significantly change the number of c-fos$^+$ nuclei in ORB (344 versus 325 c-fos$^+$ nuclei, $\chi_1^2=1$, $P=0.32$; Fig. 7i). We instead found that cocaine significantly increased the c-fos intensity levels ($t_{492}=15.21$, $P<0.001$; Fig. 7j), as well as the c-fos signal variance (Bartlett's $K^2=174.21$, degrees of freedom = 2, $P<0.001$; Fig. 7k).

To further map the brain regions that are functionally linked to increased locomotion after cocaine administration, we used linear mediational analysis on the whole-brain c-fos data. We first characterized the marginal distribution of the behavior across animals by a monotonic four-parameter cumulative Weibull distribution (Fig. 7l). As a measurement of the relative contribution of a specific brain region to the mediated effect, we estimated the proportion of the total effect of cocaine on locomotion that is mediated by c-fos expression (i.e., relative mediated effect; Fig. 7m). This analysis revealed that ORB is a key region mediating the behavioral effect and accounted for approximately 7% of the relative mediated effect ($P_{\alpha\beta}=0.071$, 95% confidence interval: [0.00–0.898]). We found that the c-Fos fluorescence intensity correlated to velocity better than absolute c-fos$^+$ cell counts (Fig. 7n,o) and that variance in cell counts increased with average number of c-fos$^+$ neurons (Supplementary Fig. 11). We therefore normalized the variance in c-fos fluorescence intensity to DAPI in subsequent analyses. We asked whether we could classify the observed relation in ORB between c-fos induction and locomotion into a distinct form of gain control: input gain, output gain or range compression. (Fig. 7p–r). We kept the saturation level fixed at a velocity of 90 cm/s and estimated the fixed-effect of cocaine on the other two cumulative Weibull parameters (slope and intercept; Fig. 7s). Using the marginal cumulative density function as prior, we sampled from the posterior density distributions of the fixed effect of cocaine. We compared the posterior fixed effect of cocaine on either intercept or slope to that of our prior $N(0, 0.1)$ based the Savage–Dickey density ratio test. We found that the cocaine-induced locomotion most likely depended on modulation of the input gain in orbital cortex ORB (Bayes factor $(BF)_{10}=1.4$) rather than range compression ($BF_{10}=1.1$; Fig. 7t). Taken in combination with our tracing dataset (in which ORB was identified as a contextual layer preferentially targeting D1$^+$ striatal neurons) these data support ORB as a major corticostriatal hub mediating locomotor effects of cocaine.

Discussion

For invertebrate model organisms, there has been considerable progress in developing methods to map and visualize the complete nervous system connectome based on electron microscopy reconstructions[39–41]. Although detailed reconstruction of the mouse nervous system at a similar scale would be valuable, there are considerable challenges in acquiring and analyzing data at that scale[42]. To define the organization of circuits in the mouse brain, it is now becoming possible to generate a complete map of cell-type-specific connectivity using combinations of advanced transgenic, labeling and imaging methods at the cellular scale. In spite of great progress in labeling methods, methods for registering and mapping neurons have, to a large extent, still relied on manual inspection or registration, and have not been scalable for analysis of large datasets or for integration of multiple projects into one database. To address these needs, we have developed a computational framework to facilitate comprehensive investigations of mouse brain circuits at the cellular level.

We have developed a computational framework that enables automated segmentation of labeled neurons at a whole-brain scale, independent of imaging method. Based on this computational framework, data can be analyzed and presented in a scale-invariant whole-brain mouse atlas that allows rapid data comparison and data sharing between projects and laboratories. A key advantage of our framework is the ability to rapidly visualize and share whole-brain data using commonly available infrastructure. To facilitate integration of this platform, we provide video tutorials for common applications (http://wholebrainsoftware.org/tutorials).

We present the application of our computational framework to whole-brain mapping of rabies tracing data derived from different neuron subtypes in the corticostriatal pathway and in combination with whole-brain definition of cocaine-induced neuronal activity using c-fos mapping. Using whole-brain rabies tracing, we determined that ORB displays characteristics of a contextual layer in the corticostriatal pathway, possibly shaping interactions between primary motor cortex and the striatum, and furthermore that it preferentially targets the direct-pathway D1$^+$ striatal neurons. We could further integrate this connectivity pattern with cocaine-induced activity changes, using whole-brain segmentation of c-fos levels in single neurons after cocaine administration, which identified the ORB as a major mediator of cocaine-induced locomotion.

We anticipate that the versatility and accessibility of this framework will promote integration of new technologies and user-based modifications. For example, neuroanatomical methods that leverage the power of next-generation sequencing will be important for the development of computational neuroanatomy, including lineage tracing by barcoded DNA[43], connectivity tracing of individual neurons by expressing RNA barcodes[44], in situ sequencing[45,46] and spatial transcriptomics[47].

We expect that an increasing focus on whole-brain characterization of neurons, circuits and activity will benefit from a unified and intuitive approach to data analysis. Notably, our framework can integrate future advances in imaging, incorporate new molecular data to improve neuroanatomical definitions and support data sharing in mapping projects.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-017-0027-z.

Received: 30 January 2017; Accepted: 17 October 2017;
Published online: 4 December 2017

References

1. Alivisatos, A. P. et al. The brain activity map project and the challenge of functional connectomics. *Neuron* **74**, 970–974 (2012).
2. Mitra, P. P. The circuit architecture of whole brains at the mesoscopic scale. *Neuron* **83**, 1273–1283 (2014).
3. Denk, W., Briggman, K. L. & Helmstaedter, M. Structural neurobiology: missing link to a mechanistic understanding of neural computation. *Nat. Rev. Neurosci.* **13**, 351–358 (2012).

4. Bohland, J. W. et al. A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLOS Comput. Biol.* **5**, e1000334 (2009).
5. Oh, S. W. et al. A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
6. Zingg, B. et al. Neural networks of the mouse neocortex. *Cell* **156**, 1096–1111 (2014).
7. Osten, P. & Margrie, T. W. Mapping brain circuitry with a light microscope. *Nat. Methods* **10**, 515–523 (2013).
8. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
9. Tomer, R. et al. SPED light sheet microscopy: fast mapping of biological system structure and function. *Cell* **163**, 1796–1806 (2015).
10. Freeman, J. et al. Mapping brain activity at scale with cluster computing. *Nat. Methods* **11**, 941–950 (2014).
11. Hunnicutt, B. J. et al. A comprehensive thalamocortical projection map at the mesoscopic level. *Nat. Neurosci.* **17**, 1276–1285 (2014).
12. Hintiryan, H. et al. The mouse cortico-striatal projectome. *Nat. Neurosci.* **19**, 1100–1114 (2016).
13. Renier, N. et al. Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* **165**, 1789–1802 (2016).
14. Ye, L. et al. Wiring and molecular features of prefrontal ensembles representing distinct experiences. *Cell* **165**, 1776–1788 (2016).
15. Kim, Y. et al. Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Rep.* **10**, 292–305 (2015).
16. R Development Core Team. *R: a Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2008).
17. Ahrens, M. B. et al. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–477 (2012).
18. Wickersham, I. R. et al. Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).
19. Ng, L. et al. An anatomic gene expression atlas of the adult mouse brain. *Nat. Neurosci.* **12**, 356–362 (2009).
20. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
21. Mallat, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE* Trans. *Pattern Anal. Mach. Intell.* **11**, 674–693 (1989).
22. Suzuki, S. & Abe, K. Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.* **30**, 32–46 (1985).
23. Friston, K. J., Penny, W. D., Ashburner, J., Kiebel, S. J. & Nichols, T. E. *Statistical Parametric Mapping: the Analysis of Functional Brain Images*. (Academic Press, London, 2006).
24. Johnson, G. A. et al. Waxholm space: an image-based reference for coordinating mouse brain research. *Neuroimage* **53**, 365–372 (2010).
25. Allen Mouse Common Coordinate Framework. Technical white paper: Allen mouse common coordinate framework. http://help.brain-map.org/download/attachments/2818171/MouseCCF.pdf (2015).
26. Mitra, J. et al. A thin-plate spline based multimodal prostate registration with optimal correspondences. *Signal-Image Technol. Internet-Based Syst. SITIS 2010 Sixth Int. Conf.* 7–11 (2010).
27. van Kaick, O., Hamarneh, G., Zhang, H. & Wighton, P. Contour correspondence via ant colony optimization. in *Proc. 15th Pacific Conference on Computer Graphics and Applications (PG'2007)* 271–280 (2007).
28. Bookstein, F. L. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE* Trans. *Pattern Anal. Mach. Intell.* **11**, 567–585 (1989).
29. Quan, T. et al. NeuroGPS-Tree: automatic reconstruction of large-scale neuronal populations with dense neurites. *Nat. Methods* **13**, 51–54 (2016).
30. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74**, 858–873 (2012).
31. Pollak Dorocic, I. et al. A whole-brain atlas of inputs to serotonergic neurons of the dorsal and median raphe nuclei. *Neuron* **83**, 663–678 (2014).
32. Jordan, M. I. & Jacobs, R. A. Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.* **6**, 181–214 (1994).
33. Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* **70**, 119–136 (1998).
34. Wall, N. R., De La Parra, M., Callaway, E. M. & Kreitzer, A. C. Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron* **79**, 347–360 (2013).
35. Guo, Q. et al. Whole-brain mapping of inputs to projection neurons and cholinergic interneurons in the dorsal striatum. *PLoS One* **10**, e0123381 (2015).
36. Sagar, S. M., Sharp, F. R. & Curran, T. Expression of c-fos protein in brain: metabolic mapping at the cellular level. *Science* **240**, 1328–1331 (1988).
37. Graybiel, A. M., Moratalla, R. & Robertson, H. A. Amphetamine and cocaine induce drug-specific activation of the c-fos gene in striosome-matrix compartments and limbic subdivisions of the striatum. *Proc. Natl. Acad. Sci. USA* **87**, 6912–6916 (1990).
38. Pich, E. M. et al. Common neural substrates for the addictive properties of nicotine and cocaine. *Science* **275**, 83–86 (1997).
39. Zheng, Z. et al. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. Preprint at *bioRxiv* https://doi.org/10.1101/140905 (2017).
40. Eichler, K. et al. The complete connectome of a learning and memory centre in an insect brain. *Nature* **548**, 175–182 (2017).
41. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil.* Trans. *R. Soc. Lond. B* **314**, 1–340 (1986).
42. Helmstaedter, M. Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nat. Methods* **10**, 501–507 (2013).
43. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
44. Kebschull, J. M. et al. High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* **91**, 975–987 (2016).
45. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
46. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010).
47. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

Acknowledgements

D.F. thanks J. Bernardi and J.H. Lee for fruitful discussions about the software or the paper. K.M. acknowledges financial support from the Swedish Research Council (VR 2012-02049), from the Karolinska Institutet (KID-funding supporting D.F., O.T., A.M.), from the Strategic Neuroscience Area at Karolinska Institutet (StratNeuro) for rabies virus production and from the Swedish Brain Foundation (Hjärnfonden). Additional financial support for the project was from a National Institute of Mental Health grant (MH109795 to G.R., K.M. and C.A.M.) and a National Institute on Drug Abuse grant (DA0036376, to C.A.M.).

Author contributions

D.F. conceived and developed the computational framework, performed experiments and contributed to data collection, analyzed data and wrote the paper. T.V. contributed to the computational framework, performed experiments, collected data and contributed to writing of the paper. O.T., Y.X., A.M., I.L. and G.S. performed experiments and contributed to data collection. G.F. supervised experiments on corticostriatal tracing. R.T. and K.D. developed and performed COLM experiments. M.C. supervised experiments and contributed to writing of the paper. C.A.M. supervised experiments and contributed to the computational framework and to writing of the paper. G.R. supervised experiments and contributed to the computational framework and to writing of the paper. K.M. conceived and supervised the project and wrote the paper.

Competing interests

D.F. is a stakeholder in Histohub AB. The other authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-017-0027-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Virus construction and production. A protocol modified from Wickersham, Sullivan and Seung[48] was used to produce rabies virus. To make EnvA pseudotyped rabies virus, un-pseudotyped rabies viral particles were transferred onto the BHK-EnvA cell line on day 1. To decrease the contamination of glycoprotein coated rabies virus, cells were then trypsinized, washed and replated in a new cell culture flask. This procedure was repeated for 2 d. Collected supernatants were then filtered by 0.45-µm filter and ultracentrifuged at 60,000 *g* for at least 2 h on day 4. TVA, mCherry and RG were amplified, respectively, from pCMMPTVA800[18] (Addgene plasmid 15778), pAAV-EF1a double-*loxP*-flanked-hChR2(H134R)-mCherry-WPRE-HGHpA (Addgene plasmid 20297) and pHCMV-Rabies G[49] (Addgene plasmid 15785) by Phusion High-Fidelity DNA polymerase (Finnzymes, Finland). One round of overlapping polymerase chain reaction (PCR) was then performed to fuse TVA and mCherry together, using TVA Forward (NheI) and mCherry Reverse (AscI) as primers. A similar approach was used to create a TVA-t2A-RG PCR product. Inserts were digested by NheI and AscI (New England Biolabs, UK) and cloned into the pAAV-EF1a-DIO plasmid backbone. Transformation was done in NEB-10b *Escherichia coli* (New England Biolabs, UK), and bacteria were plated on 50 mg/mL ampicillin LB agar plates. Plasmid DNA was extracted from cultured colonies and confirmed by NheI and AscI digestion. Positive clones were sent for sequencing (Eurofins MWG Operon, Germany), and the results were analyzed with National Center for Biotechnology Information (NCBI) BLAST. AAV production from plasmids was done by the Gene Therapy Center Vector Core at the University of North Carolina (North Carolina, USA).

Viral injections. Animal experiments were carried out following guidelines of the Stockholm municipal committee, the Scripps Research Institutional Animal Care and Use Committee and with the approval of the Stanford University Institutional Review Panel. All viral injections into the striatum (Drd1a-cre (EY262)[50], $n=4$; Drd2-cre (ER44)[50], $n=3$; Adora2a-cre (GENSAT: KG139), $n=1$; Chat-IRES-cre (Jackson Labs: 006410), $n=4$) were done at the following coordinates from bregma: AP, +0.6 mm; ML, +2.2 mm; DV, −3.5 to −3.2 mm (500 nL). Injections into the motor cortex (Camk2a-cre (Jackson Labs: 005359), $n=4$; Gad2-cre (Jackson Labs: 010802), $n=4$) were done at AP 1.5 mm, ML 1.5 mm, DV −0.4 mm (500 nL). Injections were performed in 2- to 6-month-old mice under isoflurane anesthesia; the mice were injected with buprenorphine after surgery. In all mice, the second injection was performed after 2–4 weeks with (EnvA)SADΔG-EGFP or (EnvA) SADΔG-ChR2(H134R)-mCherry (Adora2a-cre, $n=1$; D1-cre, $n=2$; Chat-cre, $n=3$) at the same location. For the striatal injections we either used AAV5-TVA-t2A-RG or a dual-vector helper-virus system (AAV5-EF1a-DIO-TVA-mCherry and AAV5-EF1a-DIO-RG at a 1:1 ratio). In experiments with Camk2a-cre and Gad2-cre animals we used AAV9-CAGFLEX-RG (Addgene plasmid 48333) and AAV9-CAG-FLEX-TVA-mCherry (Addgene plasmid 48332) at a 1:1 ratio.

Immunohistochemistry. All immunohistochemistry was performed in either wild type C57BL/6J mice ($n=4$ cocaine treated, $n=4$ saline treated) for c-fos, or as in the example provided in quantitative single cell co-expression analysis provided a single Lhx6-EGFP mouse (Jackson Labs: 3839374). All animals were male and between 8 and 13 weeks old. Brain sections were cut at a thickness of 40 µm on a vibratome (Leica VT1000, Leica Microsystems GmbH) and placed in sodium citrate (10 mM sodium citrate, 0.05% Tween 20, pH 6) for 1–2 min for antigen retrieval, followed by a wash in TBST (0.3% Triton X-100). After antigen retrieval blocking was performed with 5% donkey serum in TBST for 1 h at room temperature (RT). Sections were then incubated with primary antibody—c-fos (1:1,000, Santa Cruz, anti-goat, sc-52, Lot G2612), parvalbumin (1:1,000, anti-guinea pig, Swant, Lot GP72), neuropeptide Y (1:500, anti-rabbit, Peninsula, lot T-4070) or Forkhead box protein P2 (FoxP2; 1:500, anti-rabbit, Abcam, Lot GR91556-1)—on a shaker at RT overnight. (For FoxP2 staining, $n=3$ C57BL/6J mice were used.)

The following day, sections were washed at RT with TBST for 10 min followed by incubation for 5 h at RT on a shaker with a secondary antibody at 1:500: Alexa Fluor-488-conjugated antibody (Jackson ImmunoResearch, 711-545-152, Lot: 100709), Cy5-conjugated (Jackson ImmunoResearch, 706-175-148, Lot: 110663) or Cy3-conjugated (Jackson ImmunoResearch, 711-165-152, Lot: 114518). Washing was performed at RT (20 °C) in TBST for 10 min followed with TBS (10 min) and then with 1× PBS (10 min). For c-fos staining, DAPI (1:50,000, Biotium, 40011, Lot: 11D1017) was added to PBS and incubated for 4 min at RT. Sections were mounted with SlowFade Gold antifade reagent (Life Technologies, 10 mL, S36936) and then imaged.

In situ hybridization. RNA in situ hybridization was performed using the RNAscope fluorescent multiplex assay (Advanced Cell Diagnostics, ACD, Hayward, CA) according to the manufacturer's instructions. In brief, C57BL/6J mouse brains were perfused with 4% paraformaldehyde and left for postfixation in 4%PFA overnight. The brains were subsequently cryoprotected by immersion in a 15% sucrose solution (in PBS) overnight at 4 °C, and the process was repeated over the following night with a 30% sucrose solution. Brains were then frozen in OCT on dry ice and stored at −80 °C. Cryosections (14 µm) were made using CryoStar NX70 cryostat (ThermoScientific) and stored at −80 °C until further processing.

Immediately before RNA in situ hybridization, cryosections were washed once in PBS (1×). Subsequently, the sections were boiled in pretreatment reagent 2 for 5 min, washed in ddH2O and immersed in 100% ethanol. Sections were dried at room temperature and a hydrophobic barrier was drawn around the individual sections using an ImmEdge Hydrophobic Barrier Pen (Vector Labs, Inc.). All following incubation steps were performed in a HybEZTM Hybridization System oven (ACD). Next, the sections were incubated with protease 4 solution (ACD) for 30 min at 40 °C. Then sections were washed twice in fresh ddH2O and subsequently hybridized with multiplexed probes: MmSlc32a1-Ch1 (Alexa Fluor-488), Mm-Slc17a6-Ch3 (Atto 647) and Mm-Pdyn-Ch2 (Atto 550) for 2 h at 40 °C. After hybridization, the sections were washed twice in RNAscope wash buffer (ACD), and four consecutive amplification steps were performed using the RNAscope Fluorescent Multiplex Detection reagents 1–4 (amplifier 4 variant A was used for fluorescent labeling). After the last amplification step, the sections were immersed in DAPI for 30 s and immediately covered with a coverslip, using a protective fluorescent mounting medium. smFISH experiments were imaged on confocal microscope with 40× optical magnification and fifteen optical *z*-planes.

smFISH quantification. Segmentation of individual transcripts was done on each channel separately. Before binary segmentation of transcripts the raw confocal image was processed with a gray morphology top-hat filter to decrease background noise. To assign individual transcripts to individual cell nuclei, the cell nuclei (DAPI) were segmented by first preprocessing the DAPI image by extracting the tensor structure energy, followed by binary segmentation and watershedding. The perinuclear zone was then obtained by iterative dilatations on the binary-segmented cell nuclei. Cell nuclei were then assigned to perinuclear zones by a point-in-polygon algorithm. The transcript gene identity, position within perinuclear zone and position within a given cell nuclei were used as input to the affinity propagation clustering algorithm[42], using a negative-squared-distances similarity measure and with the initial number of clusters set to the number of detected cell nuclei using the apclusterK() function in the apcluster R package[51].

Widefield imaging. Imaging was done at 10×0.40 NA (15506285 HCX PL APO 10×/0.40 CS, Leica Microsystems GmbH) on a fluorescent microscope (Leica DM6000B, Leica Microsystems GmbH) with a motorized stage automatically controlled through µManager[52]. Images were acquired by a Hamamatsu OrcaFLASH 4.0 digital camera (C11440-22CU ORCA-Flash4.0 V2 Digital CMOS camera, Hamamatsu Photonics K.K.) at 16 bit depth resolution with 2,048×2,048 pixels. All widefield imaging, except for rabies injections in Camk2a-cre and Gad2-cre animals, was done on a Leica DM6000B microscope. For Camk2a-cre and Gad2-cre mice, we used an IN Cell Analyzer 6000 with a 10×0.45 NA objective (GE Healthcare Life Sciences, Chicago, US).

Confocal imaging. Confocal imaging was done on a Zeiss LSM 510 or a LSM 800 confocal microscope (Zeiss GmbH). For co-localization with immunohistochemistry (Fig. 5) a 15-step *z*-stack covering the 40-µm section was imaged using an LSM 800 with 20×0.8 NA objective. For in situ hybridization, imaging was done using an LSM 800 with 40×1.3 NA oil immersion objective. All images had 16-bit depths. We used a Zeiss LSM 510 to acquire the confocal images in Supplementary Fig. 2. Objectives for the Zeiss LSM 510 are specified in the Supplementary Fig. 2 legend.

Light-sheet CLARITY imaging. Brains were perfused and incubated with CLARITY monomer solution, containing 1% acrylamide, 0.0125% bis-acrylamide and 4% PFA, and then polymerized at 37 °C for 6–7 h. Brains were passively cleared in SDS Borate Buffer (pH 8.5) at 37 °C for 4–5 weeks, equilibrated in Focusclear for imaging and imaged using COLM methods[53]. Analysis of light-sheet data differs from that of sectioned tissue in that the distance between adjacent optical planes is normally much smaller than the distance between adjacent sectioned planes. Therefore, an additional postprocessing step is needed for segmentation to ensure that segmented connected components that accord across *z*-planes are not counted more than once. This postprocessing step includes a connected-components algorithm running across the *z*-stacks on the binarized image.

Cocaine-induced locomotion. To measure locomotion, we used the open field test as a behavioral test for wild-type mice. Cocaine hydrochloride (C5776, Sigma-Aldrich) was dissolved in 9 mg/mL saline and administered at a single dose of 20 mg/kg by intraperitoneal injection. Mice were returned to the home cage for 5 min for the drug to begin to take effect and then placed inside the open field arena. The control group (saline) were injected intraperitoneally with the vehicle (NaCl 9 mg/mL). Mice ($n=4$ per group, all male) were monitored in open field using the TSE Multi Conditioning System (TSE Systems GmbH). Mice were randomly assigned and counterbalanced by treatment (cocaine vs. saline), time of testing (done at four time points in the middle of the day) and conditioning system (box A vs. box B). After each 20-min session, the mouse was brought back to its home cage, and 90 min after the injection brains were harvested by transcardial perfusion with 1× PBS followed by 4% PFA under sodium pentobarbital (0.15 mg/kg) anesthesia. To aid immunohistochemistry, perfusion

was done with 4% PFA in 1× PBS (pH 7.4; with picric acid and glutaraldehyde) on the day of perfusion. Brains were further fixated in 4% PFA at 4 °C overnight. All animals were housed in groups of 3–5 mice. All mice were housed on a 12:12-h nonreversed light:dark cycle.

NURBS atlas. We started from the scalable vector graphics (SVG) files of the original 2008 Allen reference atlas. We manually redefined and adjusted each plate. Next, we simplified points along the outer contours of each brain region using the same set of PCA-generating correspondence points, starting with the brain section outline. Adjacent sections were then merged in a sequential manner by the lofting technique. The NURBS atlas was adjusted manually using the trial version of Maya (Autodesk, Inc) when needed.

Computing intersection of a plane with NURBS surfaces. We use a general strategy to obtain a curve from the intersections of two surfaces, in which the steps involved are: (i) determine parametric curves, (i) transform curve segments and surface patches to Bézier form, (iii) conduct convex-hull intersection test to determine whether the curve can be approximated by a line, (iv) obtain intersection points in Euclidean space from intersecting lines and quadrilaterals and (v) connect intersection points to curves via a marching scheme. Throughout this process, a set of tests are performed at each stage, and when a proper curve cannot be approximated, the algorithm halts and throws an exception for that intersection.

JavaScript Object Notation (JSON) data format for serialization. Segmentation and registration results can either be transferred to collaborators as a tabulated tidy data frame or prepared for serialization on the web through a collection of name/value pairs based on JavaScript Object Notation (JSON). This open format describes the neuroanatomical spatial features of an image, along with nonspatial attributes. The format is called RosettaBrainJSON (http://openbrainmap.org/rosettabrainjson/) and is licensed under Creative Commons (CC BY 4.0).

R package and dependencies. Higher-level analysis and plotting functions were written in base R with image processing done through calls to C++ code via the Rcpp package[16,54]. For plotting in R of raster images, we used the png package[55]. For vector processing and import into R, we used the packages grImport[56] and XML[57]. For 3D rendering, we used the packages rgl and misc3d[58,59]. For general data handling and preprocessing we used dplyr[60]. All of these packages are dependencies of the Wholebrain R package.

The following C/C++ libraries were used: OpenCV 3.3[61] (for image processing), FFTW3 3.3[62] (for computations of the discrete Fourier transform) and Eigen 3.3[63] (for linear algebra, numerical solvers, etc.). We thank R. Hussain (wavelets; https://github.com/rafat), M. Schmieder (thin-plate splines; https://code.google.com/p/ipwithopencv) and P. Gadomski (coherent point drift; https://github.com/gadomski) for providing open source C++ classes that we built upon. C/C++ code was compiled using Apple LLVM version 8.1.0 (clang-802.0.42) on macOS, gcc 4.9 on Linux, and Rtools 3.4 with mingw-w64 v3 on Windows 7.

Multiresolution decomposition. A filter bank F transforms an input x into an output $y=F(x)$. Let G_j be the low-pass filter at the jth level and H_j the high-pass filter at the jth level. As stated in the main text, the Haar low-pass filter G_j simply averages adjacent entries of its input, in matrix form:

$$G_j=\frac{1}{2}\begin{bmatrix}1&1&&&&\\&1&1&&&\\&&1&1&&\\&&&\ddots&&\\&&&&1&1\\&&&&&1\end{bmatrix}$$

The Haar high-pass filter H_j computes half the difference between successive input samples, in matrix form:

$$H_j=\frac{1}{2}\begin{bmatrix}1&-1&&&&\\&1&-1&&&\\&&1&-1&&\\&&&\ddots&&\\&&&&1&-1\\&&&&&1\end{bmatrix}$$

where the size of the filters $n\times n$ matches the size of the image initially. Normally one would then downsample the signal to avoid antialiasing when reconstructing the signal. However, in our case it is more important to create a shift-invariant output of the filter bank, and this can be achieved by upsampling the filter by starting with the identity matrix, inserting empty rows in it and multiplying with the filter, thereby creating holes of zeros in the filter–hence the original French name *algorithm à trous* (with holes), or more commonly the stationary wavelet transform (SWT), due to its shift-invariant properties[64]. Upsampling is represented in matrix form as an array of size $2n\times n$:

$$U_j=\begin{bmatrix}1&&&&\\&1&&&\\&&1&&\\&&&\ddots&\\&&&&1\\&&&&1\end{bmatrix}$$

The combined operation of low-pass filtering and then upsampling is represented as the low-pass branch L_j:

$$L_j=U_jG_j=\frac{1}{2}\begin{bmatrix}1&1&&&&\\&1&1&&&\\&&1&1&&\\&&&\ddots&&\\&&&&1&1\\&&&&&1\end{bmatrix}$$

Similarly, the high pass branch B_j is given by:

$$B_j=U_jH_j$$

The low-pass branch is sent forward and at each stage the output from the high pass branch is stored as the set of detail coefficients d_j (Supplementary Fig. 2b). The output can hence be seen as rewriting the input x in terms of another basis to produce the output $y=F(x)$, so the output from a filter bank with three levels, $J=3$, is

$$y=\begin{bmatrix}L_3L_2L_1\\B_3L_2L_1\\B_2L_1\\B_1\end{bmatrix},x=\begin{bmatrix}a_3\\d_3\\d_2\\d_1\end{bmatrix}$$

where a_j is the set of approximation coefficients and d_j are the detail coefficients. When applying the filter bank to a two-dimensional signal, one obtains additional detail coefficients row-wise, column-wise and diagonally.

Filter banking is merely one specific way to implement hierarchical decomposition. More generally put, hierarchical decomposition via filter banks writes a signal in terms of new basis. To find this basis one can use the dilation function, also known as at the scaling function:

$$\phi(2^{j-1}t)=2\sum_k g(k)\phi(2^jt-k)$$

as well as the wavelet function:

$$w(2^{j-1}t)=2\sum_k h(k)\phi(2^jt-k)$$

For the Haar wavelet, substituting the low-pass filter into Eq. (1) gives us the case definition:

$$\phi(t)=\phi 2t+\phi(2t-1)=\begin{cases}1 & \text{if} t\in[0,1)\\0 & \text{otherwise}\end{cases}$$

The equation above gives us the Haar scaling function. Likewise the high pass filter can be substituted into Eq. (2), giving us the Haar wavelet function:

$$\phi(t)=\phi 2t-\phi(2t-1)=\begin{cases}1 & \text{if} t\in[0,0.5)\\-1 & \text{if} t\in[0.5,1)\\0 & \text{otherwise}\end{cases}$$

The Haar wavelet function is the simplest wavelet, but WholeBrain comes packaged with a total of 44 different wavelets. Which wavelet is best suited for neuroanatomical segmentation remains an open question, but we are quite satisfied with the output from Debauchies wavelet 2 (db2). In addition to the filter-bank implementation, WholeBrain comes with the possibility of constructing biorthogonal wavelets using the lifting scheme. The lifting scheme does not require any computation using the Fourier transform, and it can also be used to construct an integer wavelet transform, that is, a wavelet transform that maps integers to integers, which most of the time is more suitable for microscope images that are usually encoded with uint16-bit depths.

Segmentation of features by thresholding wavelet coefficients. Instead of running an inverse transform at each scale and hence increasing the computational time, we simply added the detail coefficients for a given scale

together by a weighted average. This weighted average was subsequently used for binary segmentation by thresholds defined by the user. Individual features such as cell bodies are segmented from the thresholded, weighted detail coefficients by a standard connected components algorithm[22]. In cases such as immunohistochemistry of cell nuclei (for example, c-fos), additional steps with the computation of tensor structure on the detail coefficients and subsequent extraction of the tensor structure trace or energy, tr(T) (see "Segmentation of processes using tensor structure," below), are performed, followed by thresholding and watershedding (Supplementary Fig. 10c,d). In the c-fos experiments we used Shannon entropy-based binary thresholding[65] for DAPI and c-fos cell nuclei segmentation.

Segmentation of processes using tensor structure. Anatomically, a cell body can more or less be perceived as a point feature; the most important parameter is its location. This is not true for a fiber tract or a cell process outgrowth, which are better conceived as geometrical features extended in space (spatially coherent) with a set of orientations at different locations. We here leverage these additional properties of coherency and orientation to increase the SNR for segmentation of fiber tracts and processes. We extracted the orientation and coherency of fluorescent signals at the set of detail coefficients where the SNR for processes peaked ($d1$), by first computing the gradient magnitude from the Scharr operator gradient[66] and then computing its associated structure tensor[67], defined for each location as the 2×2 symmetric positive matrix T:

$$T=\begin{bmatrix}\langle f_x,f_x\rangle_w & \langle f_x,f_y\rangle_w\\ \langle f_x,f_y\rangle_w & \langle f_y,f_y\rangle_w\end{bmatrix}$$

Where f_x and f_y are the Scharr derivatives of the detail coefficients $d_1(x,y)$ along the columns, x, and rows, y, respectively. Furthermore, the weighted inner product between two arbitrary images g and h is defined as:

$$\langle g,h\rangle_w=\iint_{\mathbb{R}^2} w(x,y)g(x,y)h(x,y)dxdy$$

where w is the Gaussian weighting function whose s.d., or kernel size, determines the area of interest, which is set and determined by the user's imaging resolution relative to the size of the labeled cell outgrowth processes. Once the structure tensor T is computed, it is trivial to obtain the following three properties for each location in $d1$: orientation, coherency and trace matrix.

Orientation, θ:

$$\theta=\frac{1}{2}\arctan\left(2\frac{\langle f_x,f_y\rangle_w}{\langle f_y,f_y\rangle_w-\langle f_x,f_x\rangle_w}\right)$$

Coherency, C:

$$C=\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and the smallest eigenvalue of the tensor, respectively.

Trace, tr(T), is defined as:

$$\mathrm{tr}(T)=\langle f_x,f_x\rangle_w+\langle f_y,f_y\rangle_w$$

The trace of the tensor structure is not primarily used for segmentation of processes, but, as mentioned in the "Segmentation of features by thresholding wavelet coefficients" section, it is quite useful for segmenting cell nuclei with differential fluorescent intensity (for example, immunohistochemical fluorescent labeling of immediate early genes; Fig. 6 and Supplementary Fig. 10). The coherency on a set of small-period detail coefficients, for example, $d1$, is used for binary segmentation of labeled processes.

For processes such as dendrites and axons, to obtain a thin shape that is equidistant from the boundaries of the labeled processes (i.e. a topological skeleton) we applied the Zhang–Suen thinning algorithm[68] to the binary segmented image. The skeleton is then pruned, similarly to Arganda-Carreras et al.[69], into endpoints (less than two neighboring segmented pixels), junctions (more than two neighboring segmented pixels) and slabs (exactly two neighboring segmented pixels). The endpoints and junctions, together with the segmented cell bodies, can then be seen as nodes that delineate the slabs as belonging to a subset of pixels defining vertices and all of which, in the end, are used for visualization of labeled processes online (Fig. 2c).

Since several pruning and tractography algorithms are already available[29,70–72] and the user might want to create their own, the output is also provided for each single pixel that forms the segmented topological skeletons, together with the pixel's associated orientation (represented as hue in Fig. 4e–g).

Registration with nonrigid free-form deformation (FFD) using thin-plate splines (TPS). To map the coordinates of a target image to those of a reference image, we used thin-plate splines (TPS). We start out by segmenting the overall contours of both the target brain section and the reference brain section from the standardized atlas. Using the contours of the overall brain section, as well as any detected hollow regions within the boundaries of the brain section, such as ventricles, the algorithm then extracts the principal axes using principal component analysis. Based on the intersection of the principal axes with the contour we iteratively generated a set of homologous correspondence points between the two sections by the following algorithm[26]:

1. Set level $l=1$
2. While $l\leq L$
3. Generate a midpoint q_i between p_i and p_{i+1} such that $q=\left(p_i+p_{i+1}\right)/2$, where the points are indexed based on their clockwise alignment on the contour perimeter
4. Find a point x_i on the contour between p_i and p_{i+1} such that x_i lies on the contour at the shortest distance from a line intersecting the contour and midpoint q_i, perpendicularly to the line between p_i and p_{i+1}
5. Iterate steps 3–5 until the whole contour has been traced in a clockwise order, i.e., $p_i=p_n$ and $p_{i+1}=p_1$
6. Make a new vector, p' in which all new x_i and q_i are interleaved together with p_i
7. Increment the level $l=l+1$. Set $p'=p+q+x$ as the new vector $p=p'$

The output of the above steps can be seen in Supplementary Fig. 2; the user defines the number of levels, L, needed to achieve a good correspondence. The more levels are added, the more correspondence points are generated, and the computational complexity of thin-plating itself is $O(n^3)$, where n is the number of correspondence points.

The TPS interpolation $\delta(x,y)$ minimizes the bending energy:

$$I_\delta=\iint_{\mathbb{R}^2}\left(\delta_{xx}^2+2\delta_{xy}^2+\delta_{yy}^2\right)dxdy$$

with the form:

$$\delta(x,y)=a_1+a_xx+a_yy+\sum_{i=1}^{n}w_iU\left(\left|\left|\,(x_i,y_i)-(x,y)\,\right|\right|\right),$$

where, a_1, a_x, a_y and $U(r)=r^2\log r$ are the affine parameters and w_i is the thin-plate splines parameter. The TPS interpolation should have square integrable derivatives if the following two conditions hold:

$$\sum_{i=1}^{n}w_i=0\text{and}\sum_{i=1}^{n}w_ix_i=\sum_{i=1}^{n}w_iy_i=0$$

A linear system is obtained for the TPS coefficients, and a regularization parameter, $\lambda=0.001$, is added to control amount of smoothing to avoid ripples in the obtained transform. If $\lambda=0$, an exact interpolation is performed[26,28].

Statistical analysis. For all statistical inferences made, we examined distribution assumptions with the Kolmogorov–Smirnov test of equality and Bartlett's test of equal variances. These tests are only reported when significant deviations from assumptions were observed. For nonparametric count data between two groups, we used Pearson's chi-squared test of independence. Student's t tests were used when assumptions were met, with adjusted degrees of freedom when equal variances could be assumed. Only two-sided tests were performed. For clustering of single-cell coexpression analysis, we used the R package Mclust with expected maximization and Gaussian mixture models. For linear regression analysis of locomotion regressed on c-fos intensity background, we added autofluorescence as a covariate rather than excluding data points with high autofluorescence due to staining variability. All statistical analyses were done in R[16]. For Bayesian analysis and meditational analysis, see sections "Estimating mediating effects of c-fos expression to cocaine induced locomotor activity" and "Psychophysical scaling by gradient-based Markov chain Monte Carlo (MCMC)". For the signal-to-noise ratio evaluation in Fig. 3, the Rose criteria[73] was used.

No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[30,31]. Injections considered outside of the stereotactic coordinates were excluded and not further imaged or processed. We did not perform any replication experiments. Pseudorandomization was performed for the cocaine administration study. In cocaine administration experiment, we counterbalanced the assignment of mice to each group (cocaine vs. saline) × (box A vs. box B) × (time of testing).

Estimating mediating effects of c-fos expression to cocaine induced locomotor activity. Mediational analysis was done in R using the MBESS package, following Preacher & Kelley[74].

Psychophysical scaling by gradient-based Markov chain Monte Carlo (MCMC). We assume that motor behavior observed in Fig. 7l can be described with the four-parameter Weibull density distribution, $f(x)$:

$$f(x;k;\lambda)=\begin{cases}\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}e^{-(x/\lambda)^k}, x\geq 0\\ 0, \qquad\qquad x<0\end{cases}$$

which gives the cumulative density distribution:

$$F(x;k;\lambda)=\begin{cases}1+e^{-(x/\lambda)^k}, x\geq 0\\ 0, \qquad\qquad x<0\end{cases}$$

which we directly translate into the psychophysical scaling between average velocity for the ith mouse, $\overline{v}_i$, and relative variance in c-fos activation, x_i:

$$\overline{v}_i=0.9\left(1-\exp\left(\frac{-x_i}{\lambda}\right)^k\right)+1$$

where x is defined as:

$$x=\frac{\sigma^2_{\text{c-Fos}}}{\sigma^2_{\text{c-Fos}}+\sigma^2_{\text{DAPI}}}$$

where k is the shape, or slope, parameter and λ is the location, or intercept, parameter; 0.9 and 1 are estimated from the marginal distribution of the locomotor behavior itself and set the output range. Likewise, we set $\lambda=0.55$ and $k=2.5$ based on initial fit to the marginal distribution; we then estimated the effect of cocaine with prior *Gaussian*(0, 0.1). Bayesian statistics were done in R using the rstan[75] package.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code access and supplementary software instructions. Source code can be accessed at https://www.github.com/tractatus/.

Data availability statement. The datasets generated during and/or analyzed during the current study, if not already available on http://www.wholebrainsoftware.org, http://www.github.com/tractatus or http://www.openbrainmap.org, are available from the corresponding authors on reasonable request. Source code is licensed with a GPLv3 license.

References

48. Wickersham, I. R., Sullivan, H. A. & Seung, H. S. Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high-level gene expression in neurons. *Nat. Protoc.* **5**, 595–606 (2010).
49. Sena-Esteves, M., Tebbets, J. C., Steffens, S., Crombleholme, T. & Flake, A. W. Optimized large-scale production of high titer lentivirus vector pseudotypes. *J. Virol. Methods* **122**, 131–139 (2004).
50. Gong, S. et al. Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
51. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27**, 2463–2464 (2011).
52. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using μManager. *Curr. Protoc. Mol. Biol.* **Chapter 14**, 20 (2010).
53. Tomer, R., Ye, L., Hsueh, B. & Deisseroth, K. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. *Nat. Protoc.* **9**, 1682–1697 (2014).
54. Eddelbuettel, D. et al. Rcpp: seamless R and C++ integration. *J. Stat. Softw.* https://doi.org/10.18637/jss.v040.i08 (2011).
55. Urbanek, S. png: Read and write PNG images. https://CRAN.R-project.org/package=png (2013).
56. Murrell, P. Importing vector graphics: The grImport package for R. *J. Stat. Softw.* https://doi.org/10.18637/jss.v030.i04 (2009).
57. Lang, D.T. & The CRAN Team. *XML: Tools for Parsing and Generating XML Within R and S-Plus.* https://www.rdocumentation.org/packages/XML/versions/3.98-1.9 (2017).
58. Adler, D. & Murdoch, D. rgl: 3D Visualization Using OpenGL. https://CRAN.R-project.org/package=rgl (2017).
59. Feng, D. & Tierney, L. Computing and displaying isosurfaces in R. *J. Stat. Softw.* **28** https://doi.org/10.18637/jss.v028.i01 (2008).
60. Wickham, H., Francois, R., Henry, L. & Müller, K. dplyr: a grammar of data manipulation. https://CRAN.R-project.org/package=dplyr (2017).
61. Bradski, G. The OpenCV library. *Dr. Dobbs J. Softw. Tools Prof. Program.* http://www.drdobbs.com/open-source/the-opencv-library/184404319# (2000).
62. Frigo, M. & Johnson, S. G. The Design and Implementation of FFTW3. *Proc. IEEE* **93**, 216–231 (2005).
63. Guennebaud, G. & Jacob, B. Eigen v3. (2010). Available at: http://eigen.tuxfamily.org.
64. Holschneider, M., Kronland-Martinet, R., Morlet, J. & Tchamitchian, P. A Real-time algorithm for signal analysis with the help of the wavelet transform. In: P. J.-M. Combes, P. A. Grossmann, P. P. Tchamitchian eds.. *Wavelets* (pp. 286–297. Springer Berlin, Heidelberg, Germany, 1990).
65. Huang, L.-K. & Wang, M.-J. J. Image thresholding by minimizing the measures of fuzziness. *Pattern Recognit* **28**, 41–51 (1995).
66. Scharr, H. *Optimal Operators in Digital Image Processing.* (University of Heidelberg, Germany, 2000).
67. Rezakhaniha, R. et al. Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy. *Biomech. Model. Mechanobiol.* **11**, 461–473 (2012).
68. Zhang, T. Y. & Suen, C. Y. A fast parallel algorithm for thinning digital patterns. *Commun. ACM* **27**, 236–239 (1984).
69. Arganda-Carreras, I., Fernández-González, R., Muñoz-Barrutia, A. & Ortiz-De-Solorzano, C. 3D reconstruction of histological sections: Application to mammary gland tissue. *Microsc. Res. Tech.* **73**, 1019–1029 (2010).
70. Wearne, S. L. et al. New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience* **136**, 661–680 (2005).
71. Rodriguez, A., Ehlenberger, D. B., Dickstein, D. L., Hof, P. R. & Wearne, S. L. Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. *PLoS One* **3**, e1997 (2008).
72. Wang, Y., Narayanaswamy, A., Tsai, C.-L. & Roysam, B. A broadly applicable 3-D neuron tracing method based on open-curve snake. *Neuroinformatics* **9**, 193–217 (2011).
73. Rose, A. *Vision: Human and Electronic.* (Plenum Press, New York, 1973).
74. Preacher, K. J. & Kelley, K. Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychol. Methods* **16**, 93–115 (2011).
75. Carpenter, B. et al. Stan: A probabilistic programming language. *J. Stat. Softw.* https://doi.org/10.18637/jss.v076.i01 (2017).

Corresponding author(s): Daniel Fürth (daniel.furth@ki.se)
Konstantinos Meletis (dinos.meletis@ki.se)

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications (Watabe-Uchida M, et al, Neuron 2012; Pollak Dorocic I, et al, Neuron 2014).

2. Data exclusions

Describe any data exclusions.

Injections considered outside of the stereotactic coordinates were excluded and not further imaged or processed.

3. Replication

Describe whether the experimental findings were reliably reproduced.

We did not perform any replication experiments.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Pseudo-randomization was performed for the cocaine study (Figure 7). Where wildtype animals were assigned to (cocaine vs. saline) x (boxA vs. boxB) x (time of testing), (4 vs 4) x (4 vs 4) x (4 time points).

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

There was no blinding performed.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Several parts of the method section is dedicated to the software and code is freely available on github.com/tractatus/wholebrain as well as detailed instructions for installation and use on www.wholebrainsoftware.org/

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Plasmids are available from Addgene (Addgene plasmid 15778, Addgene plasmid 20297, Addgene plasmid 15785, Addgene plasmid 48333, Addgene plasmid 48332) All mice (Jax and GENSAT) as well as reagents IHC and ISH are available from commercial vendors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	The antibodies are commonly used in the field. cfos (Santa Cruz, anti-goat, sc-52, Lot G2612) parvalbumin 1:1000 (anti-guinea pig, Swant, Lot GP72) neuropeptide- Y 1:500 (anti-rabbit, Peninsula, Lot T-4070) Forkhead box protein P2 1:500 (anti-rabbit, Abcam, Lot GR91556-1)

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	*Provide information on cell line source(s) OR state that no eukaryotic cell lines were used.*
b. Describe the method of cell line authentication used.	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used have been authenticated OR state that no eukaryotic cell lines were used.*
c. Report whether the cell lines were tested for mycoplasma contamination.	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination OR state that no eukaryotic cell lines were used.*
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	*Provide a rationale for the use of commonly misidentified cell lines OR state that no commonly misidentified cell lines were used.*

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	All mice were male and 2-6 months old. Wild-type mice were C57BL/6J (n = 9). Transgenic strains used: Camk2a-cre [Jax: 005359] n = 4, Gad2-cre [Jax: 010802], Drd1a-cre (EY262) n = 4, Drd2-cre (ER44) n = 3, Adora2a-cre [GENSAT: KG139] n = 1, Chat-IRES-cre [Jax: 006410] n = 4

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	*Provide all relevant information on human research participants, such as age, gender, genotypic information, past and current diagnosis and treatment categories, etc. OR state that the study did not involve human research participants.*