

# Identifying Neuronal Lineages of *Drosophila* by Sequence Analysis of Axon Tracts

Albert Cardona,<sup>1</sup> Stephan Saalfeld,<sup>2</sup> Ignacio Arganda,<sup>1</sup> Wayne Poreanu,<sup>3</sup> Johannes Schindelin,<sup>2</sup> and Volker Hartenstein<sup>4</sup>

<sup>1</sup>Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology, CH-8057 Zurich, Switzerland, <sup>2</sup>Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany, <sup>3</sup>Howard Hughes Medical Institute Janelia Farm, Ashburn, Virginia 20147, and <sup>4</sup>Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, Los Angeles, California 90095

The *Drosophila* brain is formed by an invariant set of lineages, each of which is derived from a unique neural stem cell (neuroblast) and forms a genetic and structural unit of the brain. The task of reconstructing brain circuitry at the level of individual neurons can be made significantly easier by assigning neurons to their respective lineages. In this article we address the automation of neuron and lineage identification. We focused on the *Drosophila* brain lineages at the larval stage when they form easily recognizable secondary axon tracts (SATs) that were previously partially characterized. We now generated an annotated digital database containing all lineage tracts reconstructed from five registered wild-type brains, at higher resolution and including some that were previously not characterized. We developed a method for SAT structural comparisons based on a dynamic programming approach akin to nucleotide sequence alignment and a machine learning classifier trained on the annotated database of reference SATs. We quantified the stereotypy of SATs by measuring the residual variability of aligned wild-type SATs. Next, we used our method for the identification of SATs within wild-type larval brains, and found it highly accurate (93–99%). The method proved highly robust for the identification of lineages in mutant brains and in brains that differed in developmental time or labeling. We describe for the first time an algorithm that quantifies neuronal projection stereotypy in the *Drosophila* brain and use the algorithm for automatic neuron and lineage recognition.

## Introduction

The access to molecular-genetic tools and the ability to reliably identify specific neurons across multiple individuals makes the *Drosophila* brain a powerful model system for dissecting neuronal function and development (Jefferis et al., 2001; Urbach and Technau, 2004; Yu et al., 2009). The nervous system of *Drosophila* and insects in general is formed by a relatively small number of genetically and structurally defined neural lineages. Each contains ~100–150 neurons that are produced by a single precursor cell, the neuroblast.

Neurons, neuronal lineages, and compartments present strong stereotypy throughout all developmental stages of *Drosophila*. Neuroblasts undergo two proliferative phases. During embryonic development, each neuroblast spawns 10–20 primary neurons, which emit axons that fasciculate with their sisters, forming cross-brain recognizable primary axon tracts (PATs) (Nassif et al., 1998). During the larval period, a second proliferative phase generates the secondary neurons (Ito and Hotta, 1992), which remain clustered and extend a cohesive bundle, the secondary axon tract (SAT) (Fig. 1) (Dumstreit et al., 2003;

Poreanu and Hartenstein, 2006). Most SATs follow the corresponding PAT across the neuropile formed by the arborizations of primary neurons (Larsen et al., 2009). Secondary neurons generally do not further differentiate until pupal stages (Nassif et al., 2003), when they extend widespread terminal arborizations which, in conjunction with remodeled arbors of surviving primary neurons, generate the adult brain neuropile.

Neuronal lineages represent genetic (Urbach and Technau, 2004) and structural (Ito et al., 1997) modules of brain organization and provide a natural neuron-grouping system. Single neurons map to a lineage by spatial pattern of its axonal tract. Mapping neurons involves computer-assisted manual labor in which expert neuroanatomists perform crucial structure recognition decisions relative to published descriptions. Such manual analysis is very laborious and requires substantial training. The practicality of annotating and mapping each of the >100,000 neurons of the adult fly brain depends on availability of semiautomated methods for improved identification speed and quantitative measurement of annotation reliability.

Our goal is to automate the task of identifying neuronal lineages in the *Drosophila* brain. We have developed the Neurite Identification Tool (NIT) that takes as input a traced three-dimensional (3D) neurite and a set of fiducial landmarks and identifies its corresponding lineage by automated comparison with a reference library of manually annotated SATs. NIT performs pairwise global sequence alignment between the input trace and each reference annotated trace and measures multiple shape- and Euclidean distance-based parameters using the correspondences. A machine learning classifier trained with manual

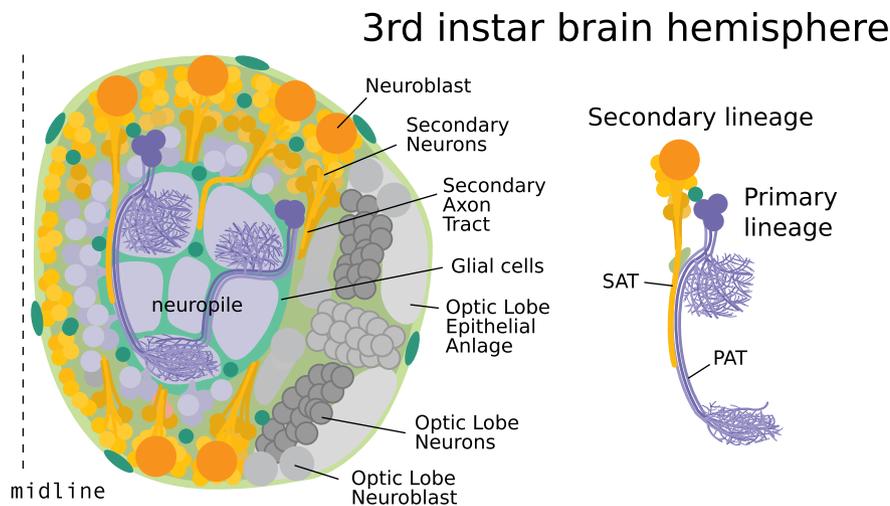
Received Jan. 12, 2010; revised March 17, 2010; accepted March 29, 2010.

This work was funded by European Union Grant 216593 "Self Constructing" (SECO) and National Institutes of Health Grant R01 NS29357-15 (to V.H.). We thank Shana Spindler and Louie Garcia for confocal image stacks of glia-less brains and dopaminergic neurons and Sergio Jiménez for help with designing and setting up a classifier with WEKA. Thanks to Parvez Ahammad and Marta Zlatić for critical comments on this manuscript.

Correspondence should be addressed to Albert Cardona, Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. E-mail: acardona@ini.phys.ethz.ch.

DOI:10.1523/JNEUROSCI.0186-10.2010

Copyright © 2010 the authors 0270-6474/10/307538-16\$15.00/0



**Figure 1.** The third instar brain and secondary neuronal lineages. Schematic diagram of a hemisphere of the *Drosophila* third instar brain (left) and its neuronal lineages (right). All neurons are unipolar. A neuroblast generates a primary neuronal lineage during late embryonic development and one or more secondary neuronal lineages at the mid larval stages. At the third instar larval stage, secondary neuronal lineages present unbranched projection bundles named SATs, which generally follow the corresponding PATs into the neuropile. SATs will extend their axonal and dendritic arborizations during the pupal period. Neuroblasts are depicted in orange; primary neuronal lineages, dark blue; secondary neuronal lineages, yellow; surface and neuropile glial cells, green; optic lobe, gray.

annotations labels each match between the query SAT and each reference SAT as correct or incorrect.

We have measured the robustness of NIT for SAT identification on mutants, on brains with shorter and longer SAT traces than the reference ones, and across developmental time, with satisfactory results. The accuracy of the recognition surpasses 99% for wild-type brains. Our algorithm can be extended to all developmental stages and can form the foundation of mapping individual neurons to their parent lineages.

## Materials and Methods

### Markers and stocks

**Fly stocks.** The five brains used for the generation of the SAT reference library were as follows: three brains from a flip-out screen with wild-type clones (hs-FLP, elav-Gal4, UAS-mCD8-GFP/FM7c; FRT42D tub-Gal80/Cyo) (Fung et al., 2009), an Oregon R carrying en-Gal4/+;UAS-cd8GFP/+ (Kumar et al., 2009), and the third instar brain on which the lineage nomenclature was established (Cha-Gal4,UAS-GFP; stock number 6793 from Bloomington *Drosophila* Stock Center) (Pereanu and Hartenstein, 2006).

**Clones.** To generate secondary lineage clones, we applied the FLP/FRT technique (Ito et al., 1997) to induce GFP-labeled clones in early larval brains, as detailed in the study by Pereanu and Hartenstein (2006) and Fung et al. (2009).

### Immunohistochemistry and histology

The larval secondary neurons were labeled with an antibody against the Neurotactin protein (BP106; Developmental Studies Hybridoma Bank) or an antibody against the *Drosophila* N-Cadherin protein (DN-Ex#8, Developmental Studies Hybridoma Bank). Apoptotic cells were labeled with anticlaved Caspase-3 antibody (Cell Signaling Technology, Cat #9661S). Glial cells were labeled with an antibody against the Repo protein (8D12; Developmental Studies Hybridoma Bank). Antibody staining of brains was performed as described by Pereanu and Hartenstein (2006). Confocal image stacks were acquired using a confocal microscope ( $\times 40$  objective; Laser sharp 2000 from Bio-Rad), with 1 or 2  $\mu\text{m}$  section intervals.

### Software

The software and documentation are available at <http://t2.ini.uzh.ch/nit/>. The NIT algorithms have been implemented in Clojure (<http://clojure.org>) as a TrakEM2 component (<http://t2.ini.uzh.ch/>

[trakem2.html](http://trakem2.html)). NIT is released under the General Public License (Free Software Foundation, <http://www.gnu.org/licenses/gpl-3.0.txt>).

Semiautomatic SAT tracing was implemented using the Simple Neurite Tracer library (Mark Longair, University of Edinburgh, Edinburgh, UK; <http://homepages.inf.ed.ac.uk/s9808248/image/tracer/>). Three-dimensional visualization was built using the 3D Viewer library (Bene Schmid, University of Wurzburg, Wurzburg, Germany; <http://www.neurofly.de>). Registration used a high-performance all-purpose registration library by Stephan Saalfeld (<http://fly.mpi-cbg.de/~saalfeld/Projects/>). All components are distributed as part of Fiji (<http://pacific.mpi-cbg.de>), a scientific image processing application based on ImageJ (<http://rsb.info.nih.gov/ij/>).

## Results

### Comparing SATs: the Neurite Identifier Tool

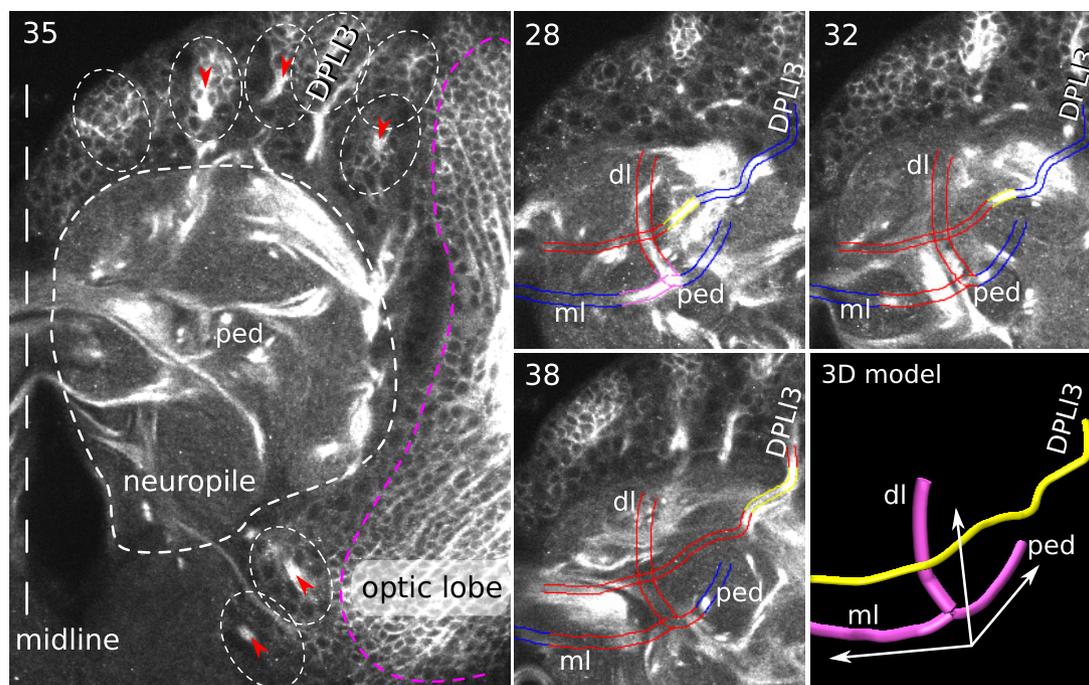
SATs are mostly unbranched at the third instar larva or present at most a single branch point (Pereanu and Hartenstein, 2006). With sparse labeling, such as with the antibody antineurotactin or with targeted GFP expression, SATs are recogniz-

able in the third instar brain as thick processes that centripetally traverse the unlabeled primary neuron cell body clusters and end up in the neuropile of primary neuron arborizations (Fig. 2).

While identifying a SAT as such is simple, the assignment of a precise identity in relation to published descriptions and nomenclature (Pereanu and Hartenstein, 2006) is extremely laborious and error prone. Currently, an experienced *Drosophila* neuroanatomist requires between 2 and 5 d to identify all secondary lineages in a neurotactin-labeled third instar brain hemisphere (2.5–6.25 lineages per hour). Particularly hard are the type II lineages (Bello et al., 2008) such as those in the centromedial (CM) and medial dorsoposterior (DPM) groups, which present anywhere between four and eight SATs each.

The automation of neuronal identification has just but started with the creation of databases for neuronal morphology, to which researchers can deposit traced neurons (Ascoli et al., 2007). Numerous methods for quantitative neuroanatomy, such as bouton spatial distribution analysis (Sholl analysis) (Sholl, 1953; Condrón, 2008), tree morphology (van Pelt et al., 1992), and estimations of synaptic density (Geinisman et al., 1996), are available. Developments in computer vision and image processing are driving automation of neurite tracing (the digitization of the three-dimensional trajectory of a neurite) and bouton and synapse recognition in light (Schmitt et al., 2004; Weaver et al., 2004; Palhares Viana et al., 2009) and electron microscopy (Macke et al., 2008; Jurrus et al., 2009; Mishchenko, 2009), substantially reducing manual labor. Quantitative analysis of neuronal structures usually take the form of analysis of dendritic and axonal arbors by tree-edit distance algorithms (the tree-edit distance is a metric to express the quantification of the number of editions necessary to transform a tree to another tree, by adding, shifting, or removing branches) (Heumann and Wittum, 2009), for the purpose of analyzing spatial distribution of synapses (Weaver et al., 2004) and neuron type classification (Ascoli et al., 2009).

There have been several attempts to identify elements of fly brains based on morphological features and to quantify their



**Figure 2.** Tracing projection bundles in confocal image stacks. View of four sections of a frontal confocal stack of an antineurotactin labeled third instar brain hemisphere. Secondary neuronal lineages are recognized as clusters of cells located in the brain cortex (dotted ellipses), presenting a central strongly labeled stalk that grows out as the SAT (red arrowheads). SATs traverse the neuropile in stereotypical patterns. Dark areas of the neuropile are composed of the arborizations of the primary neurons. Sections 28, 32, and 38 illustrate the path of a specific SAT, corresponding to secondary neuronal lineage DPLI3, relative to the mushroom body lobes [peduncle (ped), dorsal lobe (dl), and medial lobe (ml)]. Colors indicate position of the SAT trace relative to the current section: below (red), above (blue), and current (yellow). Lower right panel, 3D view of the DPLI3 SAT trace (yellow) and mushroom body (magenta).

stereotypy and variability across multiple brain samples. These studies focused on three types of structures: volumes (such as *Drosophila* brain neuropile compartments) (Jenett et al., 2006), points (such as neuronal cell bodies in the ventral nerve cord) (Bossing et al., 1996; Schmidt et al., 1997), and paths or arborizations [such as olfactory receptor neurons (ORNs)] (Brandt et al., 2005; Jefferis et al., 2007).

For approximately isometric volumes like neuropile compartments, a simple measure of their relative center of mass may suffice for identification, while the measurement of their volume and relative location may suffice for the rough quantification of their variability (Jenett et al., 2006). However, SATs are essentially linear structures winding in a 3D space, for which there is no obvious spatial center. Even small differences in the starting and ending points, and the length, may alter any center-of-mass-like measurement significantly. By their linear (rather than volumetric) nature, multiple SATs may be confined to the same enclosing volume, not differing significantly in regard to their center of mass. Some SATs may occupy nearly the exact same neuropile space and yet project in opposite directions (for example, BAMv3 and DPMpm1:2) (supplemental Fig. 1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). For this reason, precise image volume registration alone may be insufficient for identification of lineages or neuron structures in the *Drosophila* brain despite its high level of stereotypy.

One may argue that the relevance of the center of mass of a volume lies in the fact that it approximates a homologous point or region of that volume across multiple brains. Following this idea, we engaged in the search for homologous points on linear objects like SATs. For this purpose, we trace and represent SATs as a sequence of points in space and then measure relative position (shape) and absolute position (distance) properties of these sequences relative to SAT traces from reference annotated brains.

The measurement is performed on resampled SAT traces that have been aligned using their direction vector sequence representation (see Fig. 4) (see below).

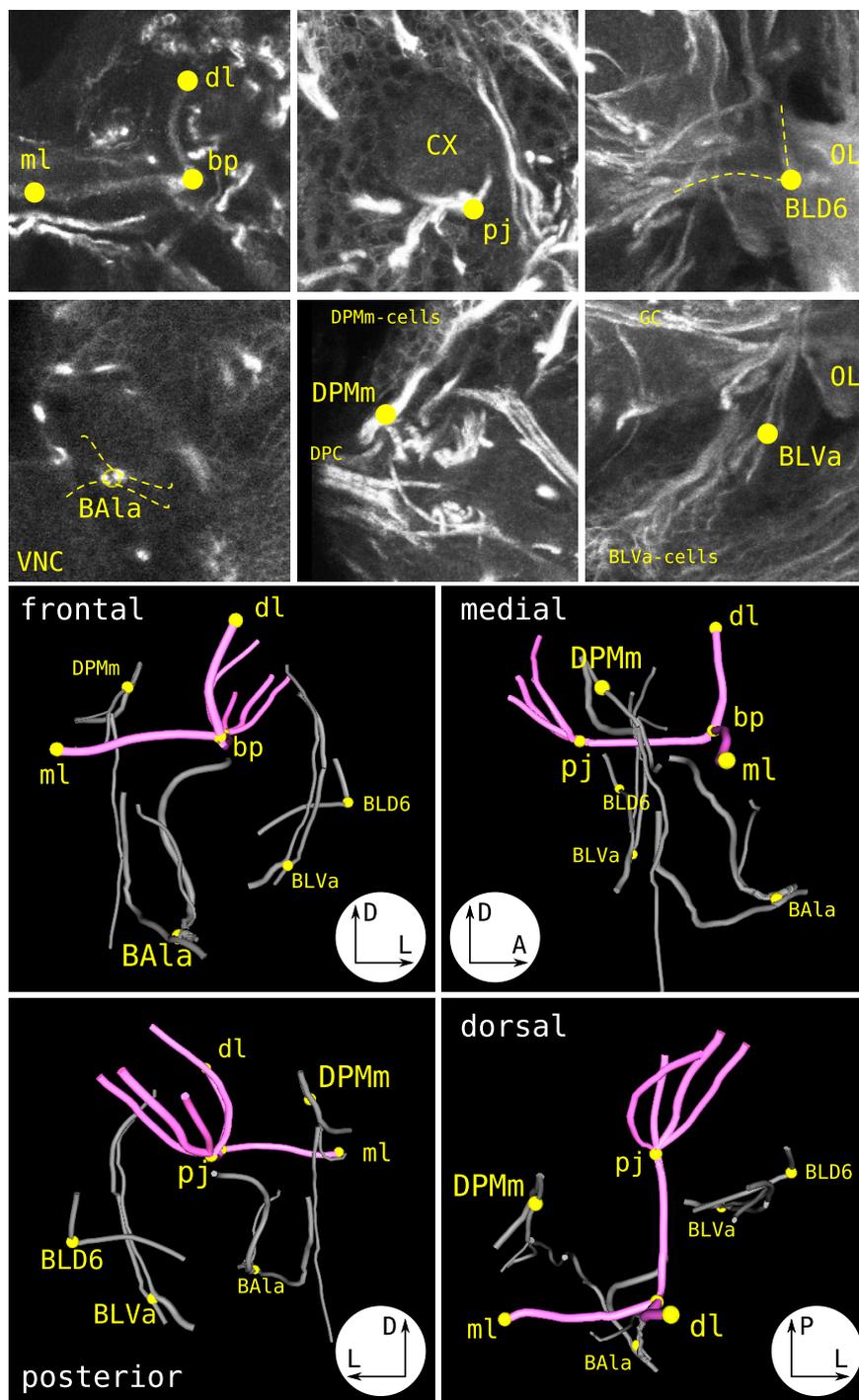
#### Tracing SATs

The first step in the identification of a SAT is its representation as a sequence of points in space. For this purpose, we have created a simple manual method: using semi-3D Bézier curves (Bézier curves whose points have a linearly interpolated z coordinate), a human operator can segment a SAT in well under a minute, approximating the skeleton of the SAT (Fig. 2). As an alternative and as part of an effort toward automation, we also use a semi-automated tracing method (Simple Neurite Tracer by Mark Longair) that automatically generates the most likely 3D path between user-defined starting and ending points as a sequence of sampled points in space.

Using these methods, we manually traced all neurotactin-labeled SATs from 5 *Drosophila* third instar brain hemispheres, building a reference database of annotated SAT traces.

#### Registering query and reference SAT traces into a common coordinate space

Each brain differs in its orientation relative to the volume defined by the confocal image stack that contains it. To compare SAT traces between brains, the latter must be registered with each other; otherwise, shape and position differences lack meaning. For this purpose, we use internal brain reference points to estimate a 3D transformation. The most obvious reference points to the *Drosophila* neuroanatomist are the mushroom body lobes, particularly the tips of the dorsal lobe, medial lobe, and peduncle, and the peduncle's branching point into dorsal and medial lobes (Fig. 3). These four reference points are easily identifiable across a variety of labelings, including no labeling (just differential background intensities).



**Figure 3.** Reference points in the brain for 3D registration. Top six panels show confocal sections illustrating the position of the eight reference points used in 3D registration. Medial to the left, dorsal to the top. dl indicates dorsal lobe tip; ml, medial lobe tip; bp, mushroom body branch point; pj, mushroom body lineages junction point, or tip of the peduncle; BLD6, sharp turn of the SAT of lineage BLD6 into the great commissure where it reaches the optic lobe (OL; dotted line indicates the SAT trace); BALa, point of divergence of the four BALa SATs (grouped as 1–2 and 3–4, their common nearby paths indicated as dashed lines; this point lies very close to the ventral nerve cord VNC); DPMm, entry point of the SAT of DPMm1 into the dorsoposterior brain commissure (DPC); BLVa, point of convergence of the three BLVa SATs, ventral to the entry of the great commissure (GC) into the optic lobe (OL). Lower four panels, 3D views of the eight reference points relative to the mushroom body lobes (magenta) and its four lineage SATs (magenta), and the SATs used as reference (gray).

The mushroom body constitutes an acceptable reference system for the following reasons. First, all lobes reasonably approximate a straight line and present clear boundaries. Second, their span is that of a large proportion of the neuropile volume, en-

par with the dimensions of the SATs to analyze. Third, each brain hemisphere contains a unique mushroom body. The mushroom bodies of the two hemispheres of a brain are mirror images of each other. Fourth, the fact that a mushroom body has three approximately perpendicular lobes naturally suggest a well posed 3D coordinate system. And fifth, the mushroom body is generally conspicuous even in the absence of a specific label, as background. Anti-neurotactin antibody, our molecular marker of choice for SATs, specifically labels the axons of newly born Kenyon cells in the mushroom body.

Our initial results suggested that the four reference points provided by the mushroom body allow for a good registration of lineages close to the mushroom body (posterodorsal, dorsal, dorsomedial, and dorsoanterior lineages), but that they are not sufficient for the lineages which are relatively far from it (the lateral, basal, and posteromedial lineages). We therefore searched for four additional reference points that are evenly spread out throughout the rest of the brain, and that are also easily recognizable in neurotactin-labeled brains. We have defined the basolateral dorsal (BLD) 6 elbow point, the basoanterior anterior-lateral (BALa) split point, the ventral basolateral (BLV) junction point, and the DPM entry point into the neuropile (Fig. 3). With all eight reference points, we estimate a nonlinear transformation using the moving least squares (MLS) method (Schaefer et al., 2006) for 3D affine transformations that provide an appropriate registration for all protocerebral and deutocerebral SATs. By this method, the brain of interest is smoothly warped onto a reference brain, preserving the relative position of internal brain components.

With the eight-point 3D MLS transformation, any query SAT can be brought with a very good approximation into the coordinate space of the reference SATs with a very fast operation. The transformation corrects for global orientation, shear, scale including mirroring (hemispheric chirality), and to a sufficient degree for local deformation (Fig. 4). We have designed our system to provide the option to constrict the desired transformation to a linear transformation [translation, rigid body, similarity (isometric scaling), and affine] or nonlinear transformation through MLS, for minimal to maximal warping, as desired.

The availability of pixel-accurate 3D registration methods (Jefferis et al., 2007) would necessarily increase the accuracy of

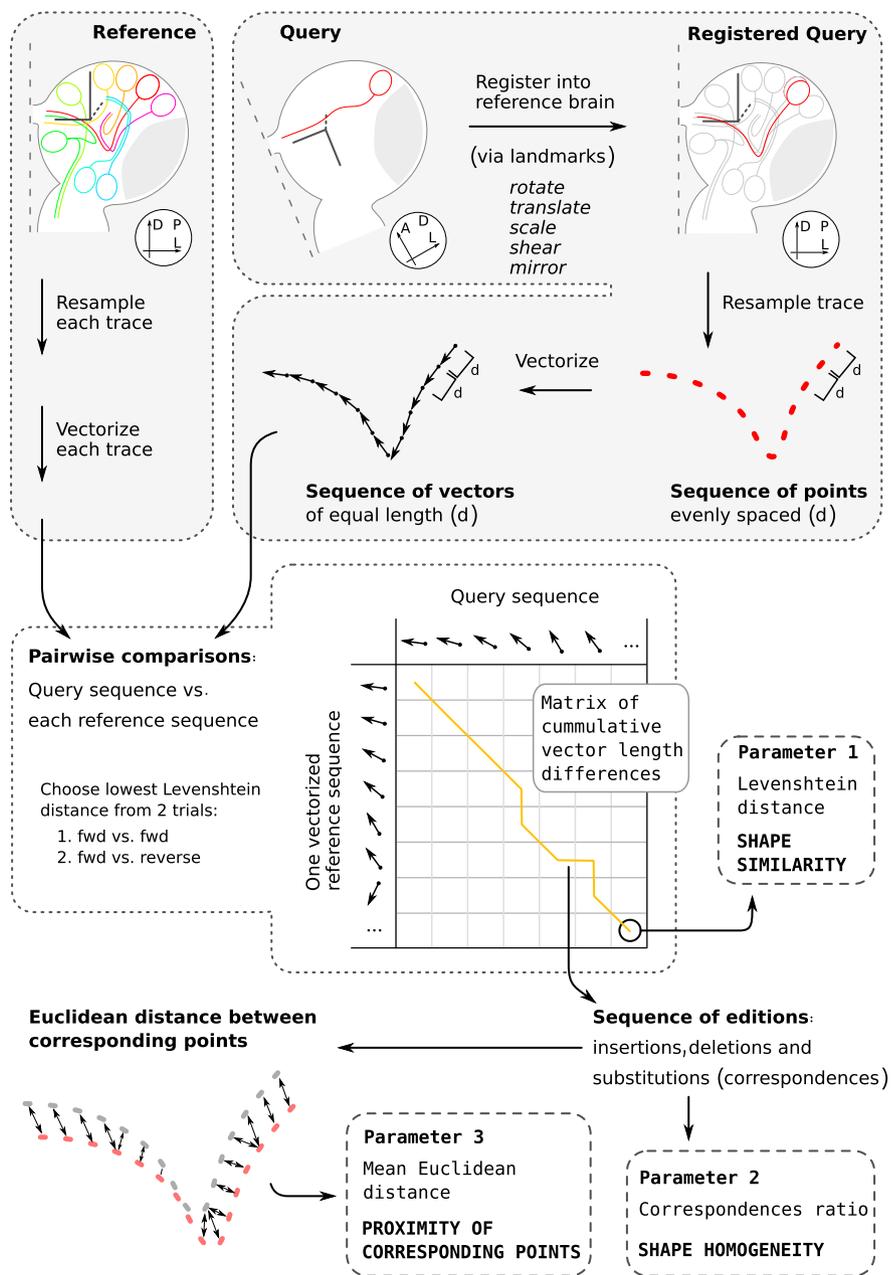
our SAT recognition system if constrained to the invariant parts of the brain. Relying on manually selected fiducial points has the practical advantages of (1) avoiding the computationally very time-consuming process of pixel-accurate brain registration, with the inevitable loss of pixel resolution, at the cost of accepting small very local inaccuracies in the registration process (we register not the brain but the traced SATs only, being a very fast computation); (2) to transition nicely into fewer fiducial points and thus less accurate brain registration (but still valuable registration), if fewer fiducial points are available due to lack of neurotactin or equivalent labeling; and (3) to improve with the availability of more fiducial points, such as obtained by automatic 3D feature extraction.

#### Comparing SAT traces

A SAT trace is a sequence of ordered points, where origin and point order matter. Jiang et al. (2002) conceived a method to generate any number of intermediate lines between any pair of lines on the plane, by applying dynamic programming to an arbitrary point sequence representation of the lines. Their method is robust as long as lines do not present loops. SAT traces are ideally suited for the application of Jiang et al. (2002) method, with modifications, for the purpose of quantifying the similarity of any pair of SAT traces. The core concept is the transformation of lines into sequences of comparable elements and then defining a cost function for the comparison of any two elements.

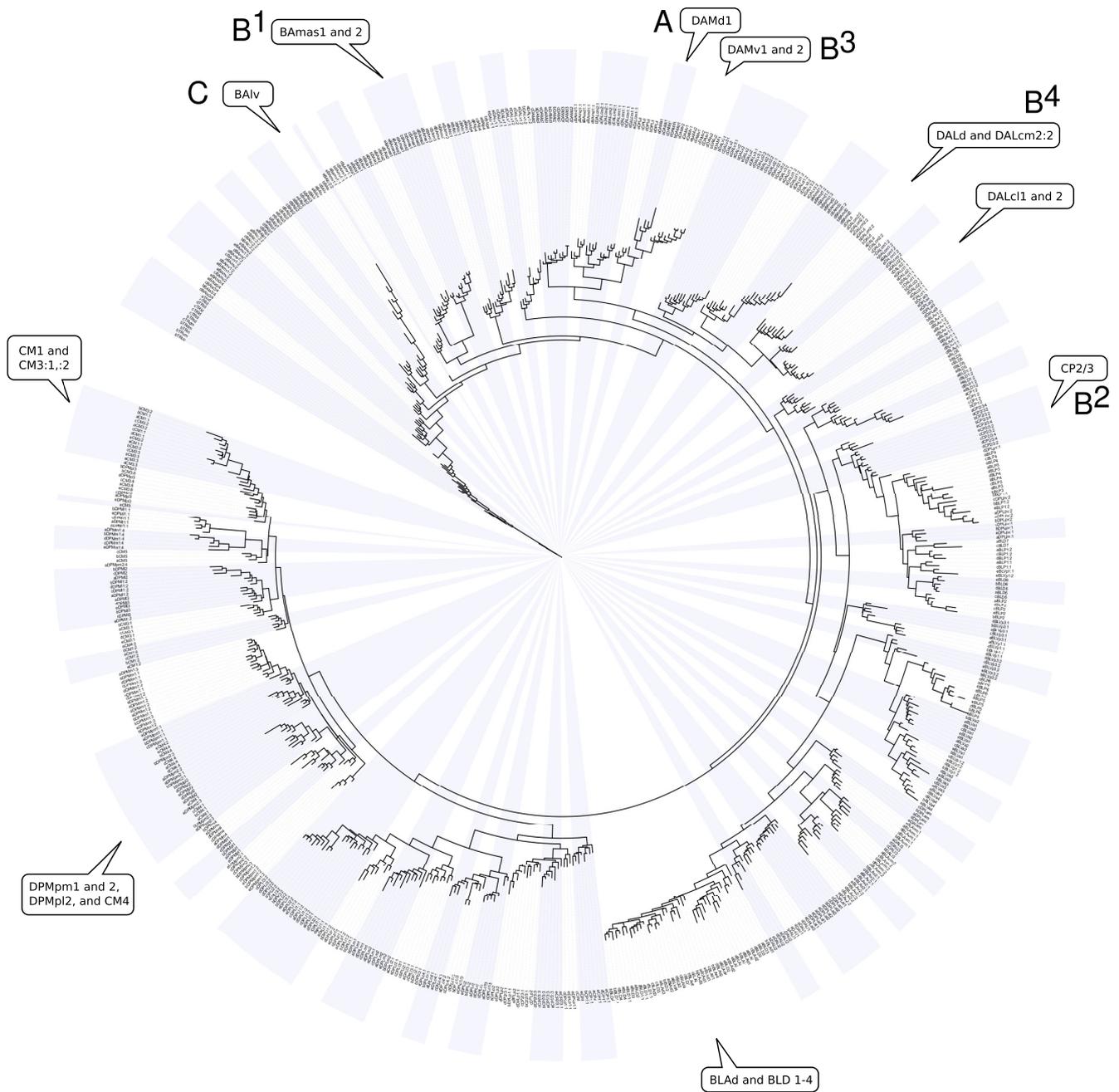
Both the manual and the semiautomatic tracing of a SAT in a confocal stack generates a sequence of points which, because of the nature of Bézier curves (or rather, their de Casteljau approximation) and the integer accuracy of pixel coordinates, respectively, present uneven consecutive point interdistances. To generate sequences of comparable elements necessary for dynamic programming comparisons, we homogenize point interdistances by resampling SAT traces with an arbitrary point interdistance  $d$ , common to all SAT traces to compare. Then, we convert the sequences of points into their corresponding sequences of direction vectors between consecutive points (Fig. 4). For any given pair of traces to compare, the number of vectors thus generated may be unequal, but the resampling to a common point interdistance  $d$  imposed an homogeneous equal vector length (Fig. 4).

The choice of the resampling point interdistance  $d$  will affect the final number of points in the sequence: when  $d$  is small, sequences will have more points and represent more accurately the original SAT trace. In the presence of various sources of noise,



**Figure 4.** The neurite search algorithm. Schematic diagram of all steps in the neurite alignment and similarity quantification algorithm. The algorithm is fed with a query SAT trace and a database of reference annotated SAT traces (top left). SAT traces (red) are schematized in the context of their third instar brain hemisphere (optic lobe in gray). Each trace is resampled to a common point interdistance  $d$  and then transformed into a sequence of direction vectors. Then the query sequence is aligned to each reference one by dynamic programming, using the subtraction of two vectors as cost function. From the resulting matrix of cumulative vector length differences, we extract three main parameters that describe the similarity of the two compared sequences (bottom). The Levenshtein distance is the value contained in the bottom-right cell of the matrix. The sequence of editions is extracted backtracing the matrix and is the basis for the measurement of the mean Euclidean distance between the corresponding points of two SAT traces (bottom left, in red and gray).

increasing  $d$  will smooth 3D traces (by eliminating high frequencies) and thereby potentially increase the overall accuracy of the recognition. The number of points, in turn, affects quadratically the number of pairwise operations to compute. The choice of an adequate  $d$  for resampling is thus critical for the optimal performance of our algorithm. The calibrated pixel size sets a lower bound (the length of the diagonal of a voxel); the diameter of the volume enclosing a trace (i.e., the brain surface) sets the upper bound. We have explored numerically a range of values for  $d$  and

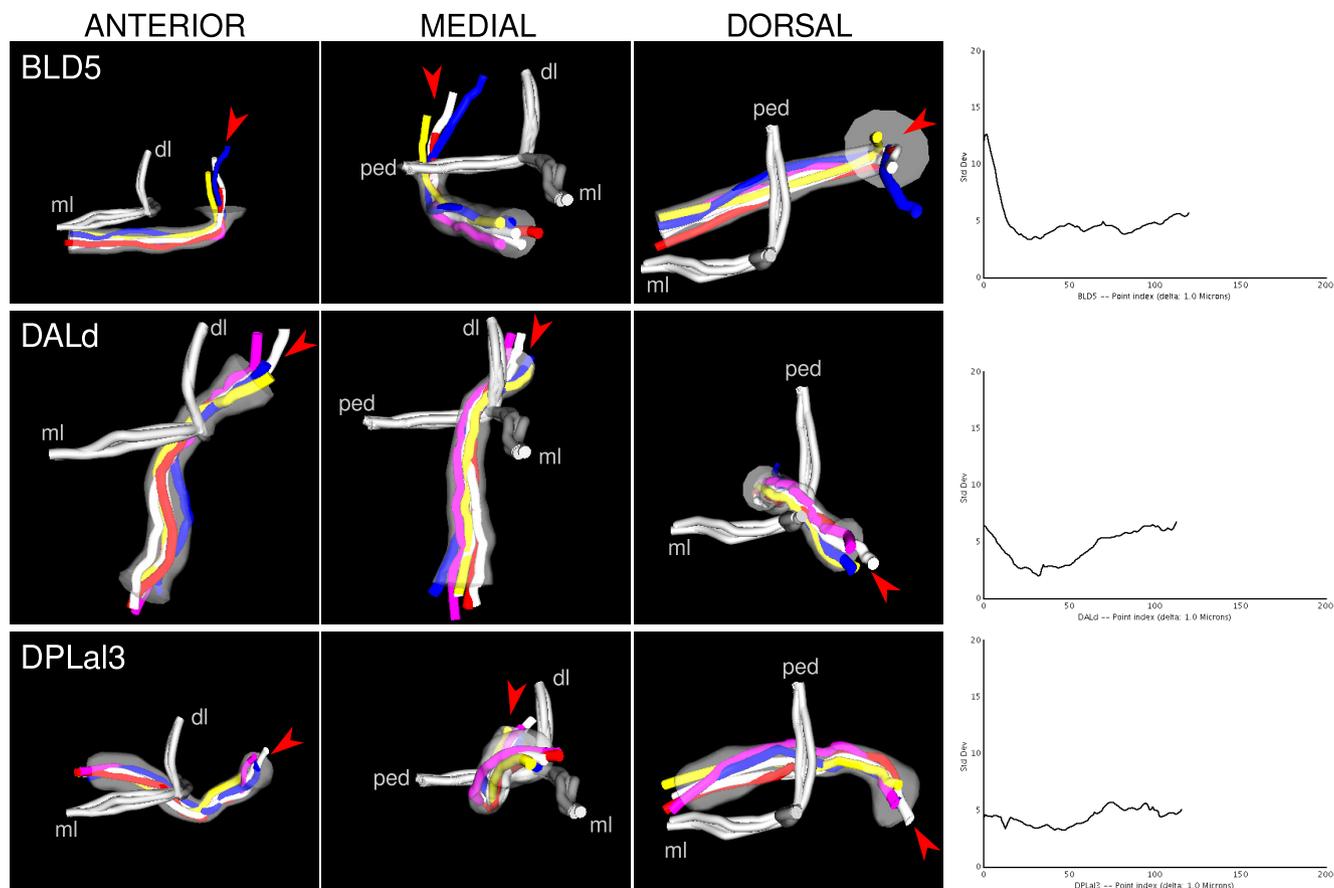


**Figure 5.** All-to-all mean Euclidean distance tree. Hierarchical clustering of all SATs of five third instar brain hemispheres based on pairwise mean Euclidean distance (MED). This figure is best viewed in its online scalar vector graphics version. The unrooted tree has been rooted at the TRco SAT for display purposes. To build the tree, we computed an all-to-all distance matrix using the MED parameter. Then we applied neighbor joining with PHYLIP (Felsenstein, 2009) and visualized the result with iTOL (Letunic and Bork, 2007). Each SAT trace has been prefixed with lowercase letters a–e to distinguish each of the five brains. If interbrain SAT trace similarity was higher than intrabrain, and the MED parameter had 100% discrimination power, we would expect all SAT traces to be clustered with their homonymous SATs in other brains in a distinct branch. While we observe this condition for approximately only half of the SAT traces (for example, DAMd1, top; gray radial stripes separate groups), there are notable exceptions. BAmas 1 and 2 (top left) present higher intrabrain similarity than interbrain, as do CP2/3:2 and :3 (right) and DAMv 1 and 2 (top) to a lesser extent. A few groups of SATs cannot be discriminated effectively with MED parameter alone, such as a group formed by BLAd and BLD1–4 (bottom right); a group formed by DPMpm1 and 2, DPMpl2 and CM4 (bottom left); and a group formed by CM1 and CM3:1 and :2 (left). SAT traces like BAIv (top left), very basal and lateral, and thus near but not enclosed by the eight reference points used for 3D registration, may still cluster together but present much higher pairwise MED (hence tree branch length) than the others. Four of 634 SAT traces are outliers in the sense that, by MED parameter, they do not cluster with their homonymous SAT traces in the other brains (eBAIc2:2, eCM5, and ddPLpv:1).

found an optimal range between 1 and 6  $\mu\text{m}$  for our *Drosophila* third instar brain confocal stacks (supplemental Fig. 2, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

After resampling SAT traces, these sequences of direction vectors constitute sequences of comparable elements: a vector from a query sequence may be compared to a vector of the reference sequence simply by subtraction (the cost function),

as shown by (Jiang et al., 2002) for the two-dimensional case. When equal, the length of the resulting vector is zero; when maximally unequal, the length is  $2d$ . With this simple cost function for correspondences, we can apply dynamic programming techniques to compare the shape of any two SAT traces (Fig. 4) (dynamic programming is a mathematical optimization method, where complex problems—such as se-



**Figure 6.** SAT minimal enclosing envelopes and variability plots. Dorsal, anterior, and medial 3D views of three lineages, BLD5, DALd, and DPLa3. Each panel shows the SAT traces of the homonymous lineages from the five reference brains, wrapped by a semitransparent minimal enclosing envelope (for detail on how the envelope is derived, see text). The mushroom body is provided as reference in gray. Red arrowhead indicates the proximal end. Plots to the right show the standard deviation of corresponding points across the five SAT traces relative to the consensus trace. The standard deviation, which correlates with the thickness of the minimal enclosing envelope, is highest at the proximal end of each SAT; it drops sharply where the SAT enters the neuropile and then typically rises slowly toward the distal end of the SAT. Supplemental Figures 4–15 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) contain a row for each SAT in the third instar brain hemisphere, grouped by lineage families.

quence alignment—are simplified by breaking them down into simpler subproblems such as the alignment of short subsequences, recursively, so that ultimately individual elements are compared). Dynamic programming requires two more costs: that of deleting and that of inserting a vector in the sequence. The trivial choice taken in the study by Jiang et al. (2002) assigns a cost of  $1d$  to both; our numerical parameter exploration suggests that a cost of  $1.1d$  is a better choice empirically (Fig. 2).

*Scoring two SAT traces for similarity: Levenshtein distance, ratio of correspondences, and mean Euclidean distance between correspondences*

Once both query and reference SAT traces have been transformed into a common reference space, resampled to an arbitrary and equal point interdistance  $d$ , and reformulated as sequences of direction vectors, we perform global sequence alignment using dynamic programming. Sequence alignment algorithms have been used previously for the alignment of sequences of discrete variables, such as for suggesting correct word spellings (Wagner and Fischer, 1974) and for nucleotide sequence alignment (Smith and Waterman, 1981; Altschul et al., 1990), as well as for alignment of sequences of continuous variables, such as for curve morphing (Jiang et al., 2002) and for recognition of *Drosophila* flight trajectory motifs (Grover and Tavaré, 2009). Inspired by Jiang et al. (2002), we formu-

lated an alignment algorithm that provides the basis for the quantification of SAT trace similarity (Fig. 4; supplemental text, available at [www.jneurosci.org](http://www.jneurosci.org)).

The global alignment matrix delivers the Levenshtein distance and the sequence of editions (the correspondences) between the elements of the direction vector sequences representing the SAT traces (Fig. 4). The Levenshtein distance measures shape similarity. However, shape information is not enough to confer identity: two SAT traces of very similar shape but far apart in the brain would score as very similar. A good parameter describing SAT trace similarity would incorporate not just shape but also Euclidean distance. Measuring Euclidean distances between two sequences of points in space is not possible without a strategy to select which points to measure against which. The sequence of editions is one such strategy, which provides point correspondences between the query and the reference sequence based on shape similarity around any given point. Thus by combining the sequence of editions, which describes point correspondences between two SAT traces, and the sequences of points in space describing each SAT trace, we compute the mean Euclidean distance between corresponding points, synthesizing information on both shape and position in the brain (Fig. 4; supplemental text, available at [www.jneurosci.org](http://www.jneurosci.org)).

We measure a total of 11 parameters from the global sequence alignment of two SAT traces (supplemental text, available at

www.jneurosci.org). The implementation of a successful SAT trace comparison algorithm revealed a number of strengths and limitations of these 11 parameters. Of note is the sensitivity of many parameters to uneven SAT trace lengths and to large insertions or deletions, as well as their value in handling branching events; on all of these we have elaborated further in the supplemental text (available at www.jneurosci.org as supplemental material).

### Construction of a digital atlas of SATs

We developed the NIT with the aim of identifying semiautomatically secondary neuronal lineages in the *Drosophila* third instar larva brain. For this purpose, we need both a comparison algorithm as described above and a reference set of annotated traces of SATs representing all the secondary neuronal lineages.

To build the reference set, we collected five brain hemispheres from nonisogenic third instar *Drosophila melanogaster* (see Materials and Methods for details). We labeled all brains with the BP106 antineurotactin antibody and imaged them with confocal microscopy from different orientations. The subsequent five confocal image stacks were imported into the software package TrakEM2, and all their distinctive SATs were manually traced as described above. Partial SAT tracing was performed in 21 additional preparations labeled with antineurotactin or an antibody against *Drosophila* E-cadherin, containing GFP-labeled flip-out clones.

We traced all SATs from all protocerebral and deutocerebral lineages from the four new stacks and from the stacks with flip-out clones. We first identified and annotated the clones manually by visual inspection, using the preparation on which the published nomenclature is based as the gold standard (Pereanu and Hartenstein, 2006). The annotation was a very laborious iterative process. Many SATs that have highly characteristic shapes and position were easily classified into a lineage group or subgroup, or even fully identified (all BA, DAL, DAM, BLVa; DPMm1, BLD5, and BLD6). Other SATs were resolved only up to the lineage group. Numerous SATs in the reference brain were traced short of their distal ending; other SATs were mistraced (e.g., DPLd and DPMcm1, and most CM). Numerous lineage SATs were fully resolved in one of the five reference brains or as a GFP-labeled clone. We propagated partial lineage identity resolution in one of the five brains to the other four brains, until all SAT traces were assigned an identity consistent with the other brains and with the published nomenclature in Pereanu and Hartenstein (2006).

To improve the accuracy of traces and annotations, and to identify errors, we clustered all SATs from all five brains by mean Euclidean distance (Fig. 5). In the resulting tree, we expect each set of five SAT instances, one from each brain, to cluster closely and separately from the rest when correctly annotated. When this was not the case, we revised the outlier SAT traces and annotations and reran the clustering. GFP-labeled lineage clones were crucial for the resolution of numerous lineages, particularly for the type II poly-SAT lineages like in the CM and DPM groups (Bello et al., 2008). The final tree contained only four outliers (see below) (Fig. 5).

### The revised secondary neuronal lineage nomenclature

The secondary neuronal lineage nomenclature described by Pereanu and Hartenstein (2006) provided individual names for most lineages. Several lineages (e.g., DALc1/2) were described as pairs of adjacent sister lineages whose SATs were so similar and close to each other that they could not be resolved. Finally, the diversity within lineages was not further considered. Numerous

**Table 1. Fraction of correct matches using mean Euclidean distance parameter**

Top 1	82.0%
Top 2	93.8%
Top 3	97.0%
Top 4	98.2%
Top 5	99.0%

**Table 2. Reliability of SAT identification using mean Euclidean distance parameter, by lineage group**

BA	372/372	100.0%
BLA	178/198	90.4%
BLD	117/128	91.4%
BLP	122/126	96.8%
BLV	155/160	96.9%
CM	138/148	93.2%
CP	116/118	98.3%
DAL	331/332	99.7%
DAM	100/100	100.0%
DPL	402/406	99.0%
DPM	268/280	95.7%

**Table 3. Random Forest classifier confusion matrix**

	Predicted good matches	Predicted bad matches
Actual good matches	2358 (99.57%)	8 (0.000%)
Actual bad matches	10 (0.004%)	292,745 (99.99%)

**Table 4. Random Forest classifier confusion matrix for all SATs of a brain not part of the reference set**

	Predicted good matches	Predicted bad matches
Actual good matches	284 (67.30%)	166 (0.003%)
Actual bad matches	138 (32.70%)	59,338 (99.72%)

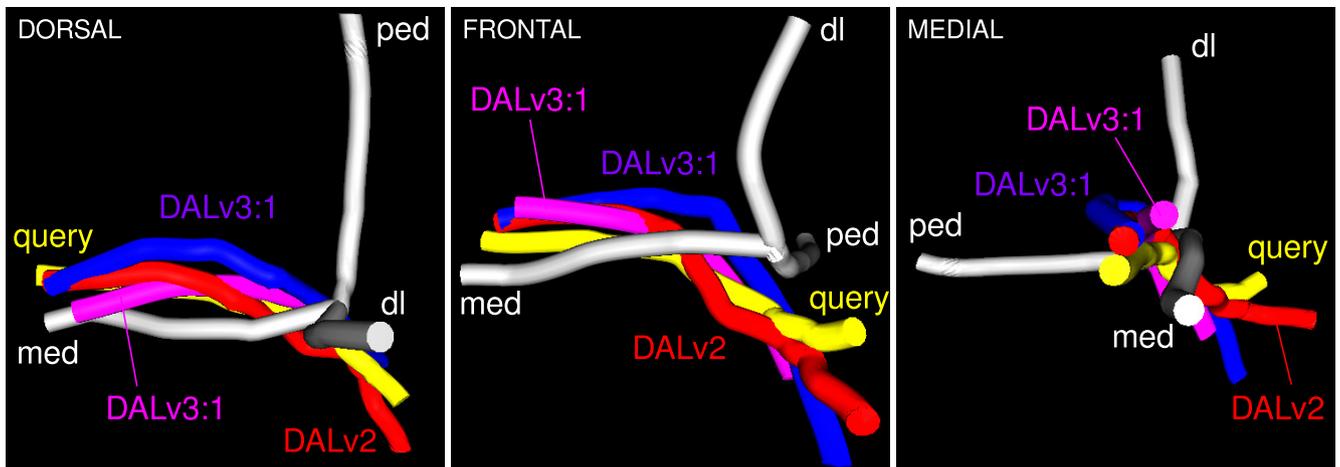
lineages are composed of smaller units [sublineages (Bello et al., 2008); hemilineages (Cornbrooks et al., 2007)] which, even though initially they project in a common SAT, they may later split into two or more branches. With the help of GFP-labeled clones, we have resolved as many as possible of the sister lineages into individual lineages and specified several SATs as sublineages.

Furthermore, we had to extend the nomenclature to distinguish individual axon tracts (SATs) formed by hemilineages or sublineages. Each SAT takes the name of the enclosing lineage plus a numerical postfix. For example, lineage CM3 has multiple prominent sublineages termed CM3:1, CM3:2, CM3:3, etc. In general, numbers are lower the more medial and dorsal the SAT trace lies.

The nomenclature described by Pereanu and Hartenstein (2006) defined a few lineages in groups consisting of adjacent or sister lineages (such as DPMp1/2; and numerous others). We have resolved as many as possible into individual lineages, with the help of flip-out clones and by comparing multiple dense-labeled neurotactin brains.

Six new lineages were added to the map published by Pereanu and Hartenstein (2006). Five of the six newly described lineages (BLD6, BLP6, DALI2, BALp4, DPMp14) could be identified in all five reference brains. BLD7 was identified in two brains only.

Four of the previously identified lineages (CM2, BALc2, DPLc5, BLVp2) turned out to be sub/hemilineages and have been removed. An updated list of all SATs is provided, which details the transition between the old and the new nomenclature (supplemental Table 1, available at www.jneurosci.org as supplemental material).



SAT	Match	Seq si...	Lev Dist	Med ...	Avg Dist	Cum ...	Std Dev	Prop ...	Prop L...	Proxi...	Prox Mut	Tortu...
DALv2 [#718 annotated]	● true	84.61	36.22	5.00	5.25	477.94	2.54	0.85	1.16	5.19	3.63	0.00
DALv2 [#380 FRT42D-WTClone]	● true	91.66	38.40	7.05	8.23	691.54	4.14	0.92	1.07	8.13	6.75	0.00
DALv3:1 [#988 #832 FRT42D-WTClone]	● true	83.33	40.92	8.65	8.61	672.00	2.78	0.84	0.84	8.50	6.63	0.00
DALv2 [#1185 L3_wayne]	● true	70.64	54.89	6.98	8.48	924.33	3.51	0.71	1.39	8.40	4.74	0.03
DALv3:1 [#660 b24shgH6R-new]	● false	83.69	50.68	9.92	10.22	940.56	3.28	0.84	1.17	10.11	7.48	0.01
DALv2 [#589 FRT42 new]	● true	85.55	43.72	9.91	9.92	893.22	4.08	0.86	1.15	9.81	7.27	0.00
DALv2 [#623 b24shgH6R-new]	● false	70.64	62.12	10.21	11.41	1243...	4.83	0.71	1.39	11.30	6.20	0.01
DALv3:1 [#1193 #1197 L3_wayne]	● false	91.66	37.13	9.95	10.52	884.05	4.48	0.92	1.07	10.40	9.02	0.01
DALv3:1 [#593 #1118 FRT42 new]	● false	74.35	45.91	12.80	12.72	992.42	5.54	0.75	0.75	12.56	10.66	0.01
DALv3:1 [#321 #713 annotated]	● false	89.53	46.71	12.12	11.79	1014...	5.36	0.90	1.10	11.65	10.43	0.02
DALcl1:2 [#1099 b24shgH6R-new]	● false	100.0	38.67	11.92	11.64	896.75	5.26	1.01	1.0	11.49	11.49	0.00
DALv3:2 [#374 #829 FRT42D-WTClone]	● false	83.33	55.91	14.95	14.34	1118...	6.87	0.84	0.84	14.16	12.29	0.01
DALv3:2 [#863 b24shgH6R-new]	● false	91.02	50.23	13.68	14.44	1126...	3.90	0.92	0.92	14.26	12.75	0.01
DALcl1:2 [#1854 L3_wayne]	● false	84.61	54.56	13.84	13.59	1237...	4.83	0.85	1.16	13.44	10.98	0.00
DALv3:2 [#1723 #1199 L3_wayne]	● false	92.77	49.09	16.61	15.61	1295...	3.03	0.93	1.06	15.42	14.25	0.01

**Figure 7.** Query of a SAT trace and visualization of matching results. An unknown query SAT traced in an antineurotactin-labeled third instar brain is compared to all SAT traces in the reference database. Results are sorted by mean Euclidean distance (MED) and classified as positive (true) or negative (false) matches. Of the five positive matches (highlighted in green), four are DALv2 and one is DALv3:1 (magenta), the latter likely a false positive. A DALv2 in the seventh row is classified as false, likely a false negative. The facts that the top two matches are DALv2, and four of five positive matches are DALv2, suggest that DALv2 is the correct annotation. In the top panels, 3D visualization of the matches with corresponding colored dots in the table. For clarity, only a true positive (red), a false positive (magenta), and a true negative (blue) are shown. The mushroom body lobes (dl, dorsal lobe; med, medial lobe; ped, peduncle) are shown for spatial reference.

The large majority of SATs could be identified in all five brains. However, a few SATs do not have homologs across all brains (see supplemental Table 2, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material), indicating either differential labeling by the antineurotactin antibody marker BP106, our inability to resolve the SAT trace in a dense neurotactin labeling (likely the case for the type II secondary lineages of the CM group; Bello et al., 2008) or variability of SATs (only two of five: BLD7, DPLc4; only three of five: BLAd3, BLD4, BLP1:1, DPMl2, DPMpm2:2; only four of five: BAMv1:1, BLAd4, CM3:2, CM3:8, CM4:1, CM5, CP1:2, DALl3, DPLl2:2, DPLpv:2, DPMpm1:1, DPMpm1:2, DPMpm2:1, DPMpm2:3).

#### Analysis of the stereotypy of SATs

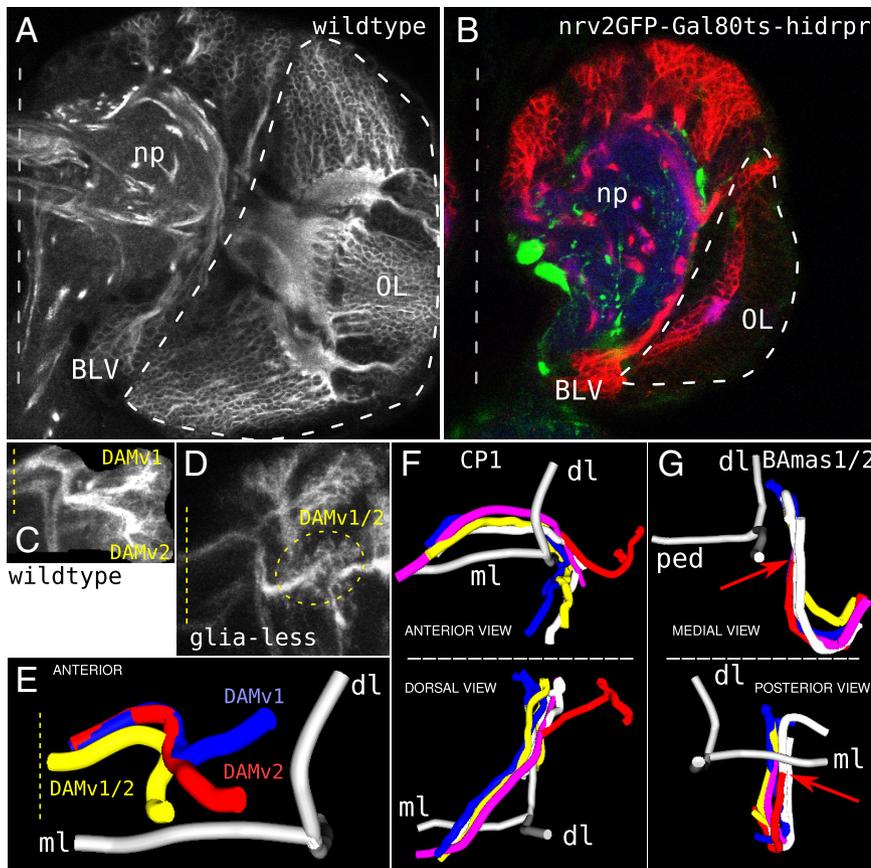
Insect brains and *Drosophila* brains in particular have been described as highly stereotypical (Hartenstein and Campos Ortega, 1997; Jenett et al., 2006; Technau, 2008). Individual cells are recognizable across brains of different individuals [for neuroblasts, see the study by Urbach and Technau (2004); for late embryo ventral nerve cord neurons, see the studies by Bossing et al. (1996) and Schmidt et al. (1997)] and in the first instar larva (personal observation). We used the most discriminative parameter of NIT, the mean Euclidean distance, to measure the inter-brain variability of SATs.

We use the tree of clustered lineages introduced above (Fig. 5). The tree illustrates that approximately half of the lineage traces

are most similar to their homonymous partners across brains and thus cluster together (for instance, DAMd1) (Fig. 5A). Only three pairs are more similar intrabrain with their sister lineages than interbrain (BAMs 1 and 2, and to a lesser degree CP2/3:2 and :3; DAMv 1 and 2; DALd and DALcm2:2) (Fig. 5B). A few sets of lineages appear unresolved. The least well resolved set consists of the BLAd and the BLD 1-4, which are all closely apposed in space, and all project into the dorsal half of the compact transverse superior fascicle (Pereanu and Hartenstein, 2006). A second unresolved set includes the CM1 and CM3:1 and :2 traces, and the DPMpm1 and 2, DPMpl2 and CM4 traces.

For the lineage BALv, all five instances (one per brain) cluster together in the same clade but present substantial distance from each other (Fig. 5C). The location of BALv, as its name indicates, is baso-anterior lateral ventral, furthest away from the eight 3D landmarks used for estimating a 3D transformation for volume registration (Fig. 3). All tritocerebral lineages are likewise affected and have thus been used as outgroups for neighbor joining.

Only 3 of 634 traces are outliers in the sense that they do not get resolved well and appear as sister branches to entire groups, near their homonymous traces in the other brains. These are e-BALc2:2, e-CM5, and D-DPLpv:1 (where the prefix “a” to “e” indicates one of the five brains). Their idiosyncratic position in the tree despite their spatial position in the brain next to nonidiosyncratic traces suggests true variability in the projection pattern of these SATs.



**Figure 8.** SAT identification in a glia-less mutant brain. **A–D**, Confocal sections. **A**, Wild-type third instar brain hemisphere labeled with antineurotactin. **B**, Third instar brain hemisphere in which neuropile glia were ablated by targeted expression of a hid-reaper construct. Observe a reduced optic lobe (OL) and neuropile (np). BLV indicates cells of the BLV lineages. **C, D**, DAMv1 and DAMv2 lineages in (**C**) wild type and (**D**) glia-less. Note that the cells of the sister lineages are fused into a single cluster in the glia-less brain. Fusion of neighboring lineages (which are normally separated by glial sheaths) is a typical phenomenon observed in glia-less brain. **E**, Wild-type DAMv1 and DAMv2 lineages (in red and blue) are compared to the glia-less fused pair (in yellow); anterior view. Despite the fusion, DAMv1 and DAMv2 lineages are recognized as DAMv1 or 2. **F**, SAT traces of CP1 lineage in wild-type and glia-less brain. Four glia-less brains were registered with one of the wild-type brains that formed part of our reference data base. The wild-type CP1 SAT trace is shown in blue; CP1 traces of glia-less brains are in white, yellow, magenta, and red. Mushroom body is shown as reference in gray. In three of four brains, CP1 SATs are unperturbed; in the fourth brain (in red), CP1 shows severe pathfinding errors. Our lineage recognition method conclusively classified all as CP1 lineage, except for the one with pathfinding errors (in red). The lack of positive matches and the elevated median Euclidean distance values of the top-scoring results suggests its condition as mutant. Its identity as CP1 was determined by elimination: all neighboring lineages have a positive match other than CP1. **G**, SAT traces of BAmas1/2 lineages wild-type (blue) and glia-less brains. Traces in three of four brains are unperturbed or slightly shorter than the wild type; in the fourth brain (red), the SAT is considerably shorter (red arrow indicates its terminal end). All BAmas1/2 SATs were identified as such using a substring global alignment approach (see supplemental text, available at [www.jneurosci.org](http://www.jneurosci.org)). The shortened terminal end was identified as a long sequence of deletions at the end of the sequence of editions resulting from the alignment. In **E–G**, the mushroom body skeletonized lobes [dorsal lobe (dl), medial lobe (ml), peduncle (ped)] are shown for spatial reference.

#### Variability along the SAT trace: minimum enclosing envelope and standard deviation

We created a consensus trace for each SAT by condensing up to five SAT traces into a single one. We measured the mean Euclidean distance for all possible pairwise combinations, and then, using an unweighted pair group method with arithmetic mean (UPGMA) (Dubes and Jain, 1988) strategy, iteratively merged the closest pair until only one trace remained. Merging was performed in a weighted manner (Jiang et al., 2002), where the weight represents the proportion of the number of original traces that each trace in the pair contributes. (E.g., in merging two original SAT traces, each contributes itself only with a weight of 1/2. In merging an original trace with a trace resulting of merging two traces, the weights are 1/3 and 2/3, respectively).

We used the consensus trace as the ideal trace of each SAT. We then measured the deviation of the real SAT traces from the ideal, in two ways: (1) by determining the minimum envelope that encloses all original SAT traces; and (2) by plotting the standard deviation at each point in the consensus SAT trace sequence. We generated the lists of points of the source traces which were used to generate each point of the consensus trace. We measured the maximum distance between the consensus point and its corresponding source points and calculated the standard deviation. As a visual representation of SAT stereotypy, we built the minimum enclosing envelope by generating a tube centered on the consensus trace and has the maximum distance as its radius at any given point. The stereotypy of a SAT is correlated to the radius of its minimum enclosing envelope.

When plotting the standard deviation of radii along the trace, we observed a pattern common to most SATs. The initial segment of the envelope (corresponding to the trajectory of the SAT as it crosses the brain cortex) has a conical shape, with the base of the cone at the brain surface. For most of the trajectory of a SAT within the neuropile, the minimum envelope is cylindrical and of relatively small diameter. Distally, toward the ending of the SAT in its target area, we typically observed another smaller widening of the envelope (Fig. 6).

#### Semiautomatic annotation of SATs

##### Reliability of SAT identification

The NIT measures 11 parameters that describe the degree of (dis) similarity between any two 3D SAT traces. We test the discriminative power of each parameter by comparing each annotated SAT of any of the five brains against the annotated SATs in each of the other four brains. For each comparison, we sorted the four sets of results and extracted from each the index of the homonymous SAT; ideally the top one. For the few combinations in

which a brain did not contain the homonymous SAT, such were not considered.

The results of the parameter analysis indicate that the mean Euclidean distance between the substitution correspondences has the highest discriminative power: 82.0% of homonymous top matches, 93.8% for an homonymous match within the top 2, and saturates after top 5 with 99.0% (Table 1). By the same parameter, the recognition of the lineage group is 97.1% (where lineage group is one of BA, BLV, etc.). All other parameters present lower discriminative power. The accuracy in the classification varies among lineage groups. Some lineage groups score better than others, ranging from 100% accuracy for SATs in groups BA and DAM to 90.4% for group BLA (Table 2).

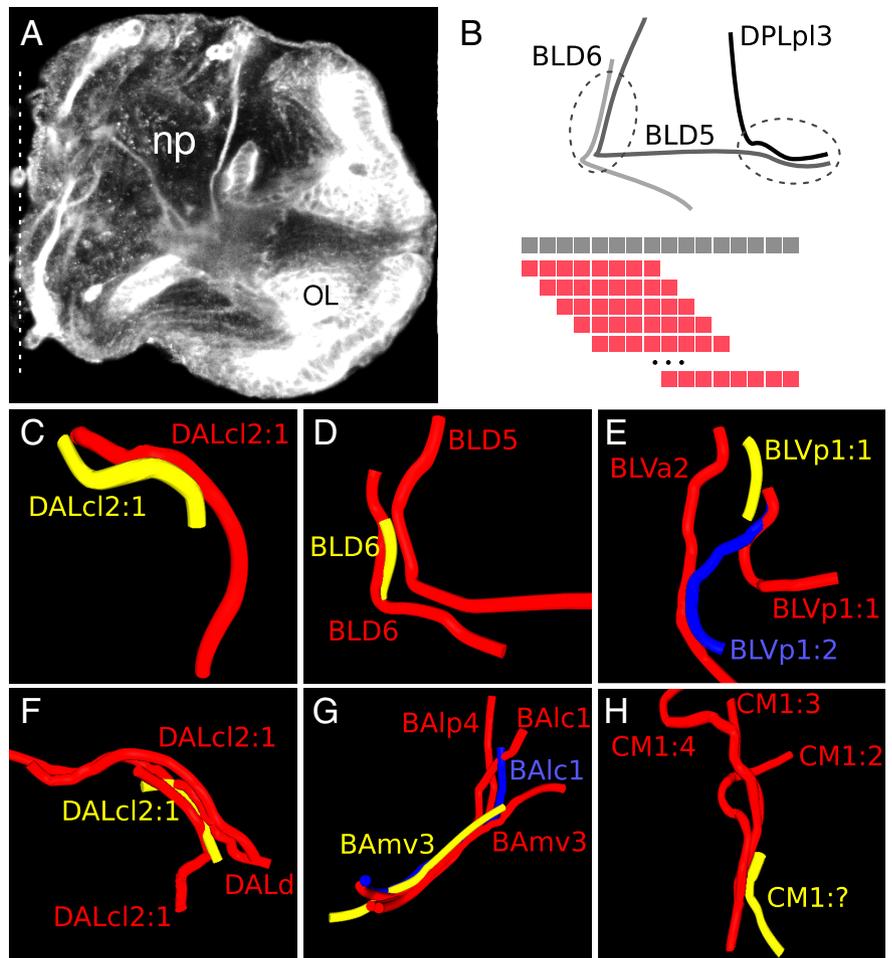
### A multiparametric classifier

The combination of multiple parameters allows for more reliable SAT identification. We use a machine learning method known as Random Forest (Breiman, 2001), which combines many decision trees (70 in our case) (supplemental Fig. 3, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material), to learn the best way of separating the samples between two different classes: correct or incorrect. The decision trees are simple binary trees in which each node divides the set of samples based on the most differentiating feature at the given tree level. In this fashion, the deeper we go in the tree, the better samples are differentiated.

We trained the algorithm with the results of comparing, on the expert-classified data, each SAT trace in one brain to all SAT traces in every other brain (five brains total, hence four lists of results for each trace). The number of expected incorrect matches is much higher than the expected correct matches; hence, to avoid overfitting for incorrect matches, we trained with only the top eight results of each test, as sorted by the most discriminative parameter (mean physical distance, which by itself never scores in our data a correct match lower than at position 8). We used an open source implementation of the Random Forest approach (WEKA library) (Witten and Frank, 2005) and stored the trained model for practical application to SAT classification.

The Random Forest approach results in 99.7% (2358 of 2368) correct good matches and 99.98% correct bad matches (292,745 of 292,753), with only eight false positives and 10 false negatives (Table 3). All eight false positives involve matches between DAMd2 and DAMd3, two extremely similar and closely overlapping SATs.

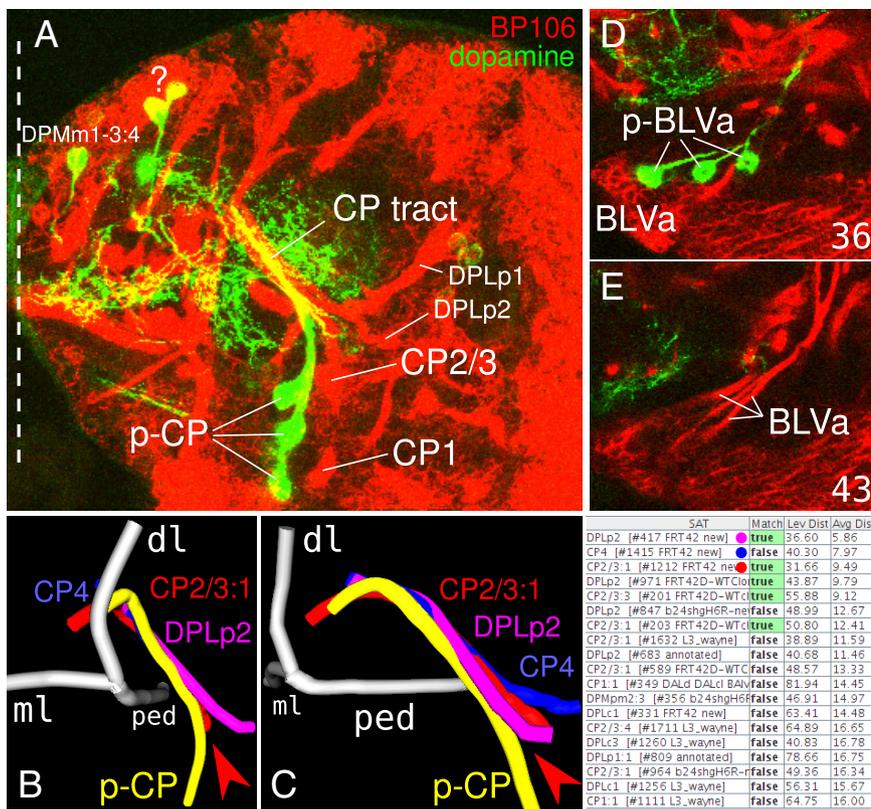
We tested the reliability of the classifier in identifying SATs of a brain not belonging to the training set. We traced 80 SATs, including the hardest lineage groups, BLA and BLD, in full. The classifier presented a high number of false positives (Table 4), but these were nonhomogeneously distributed: 42 of 80 traces had zero false positives, and another 20 had three or fewer false positives. In 75 of 80 SAT traces, the classifier found at least one true positive (true positives:  $3.30 \pm 1.66$ ; false positives:  $1.71 \pm 2.31$ ; false negatives:  $1.89 \pm 1.61$ ). The five SATs without a true positive were relatively short SATs (BALp3, BALp4, BLD1:2, BLD7, and DPLm1). When sorting results by the mean Euclidean distance, the top result was correct in 68/80 cases, and the top two results contained a homonymous SAT in 76/80 cases. An example of comparing an unknown SAT to all traced SATs in the database is shown in Figure 7.



**Figure 9.** Identifying incomplete SAT traces. **A**, Confocal section of a wild-type third instar brain hemisphere labeled with anti-DE-cadherin. Medial to the left, dorsal is up. The proximal segments of SATs are visible as thick white lines; distal segments in the neuropile are for the most part invisible due to the fact that DE-cadherin is also expressed by the surrounding neuropile glia, which masks the SATs. **B**, Identifying SATs from small fragments is difficult when the fragment is part of a common tract shared by multiple SATs. For example, in the diagram shown here, BLD6, BLD5, and DPMp13 overlap at different positions (dotted circles). We perform global sequence alignment of the short trace (red) over all possible continuous subsequences of the longer trace (gray). Hence, the two sequences being compared have the same length, avoiding sequence length-derived issues at the cost of an increased rate of false positives, while still optimizing for the best gap-less fit (see supplemental text, available at [www.jneurosci.org](http://www.jneurosci.org)). **C–H**, Three-dimensional models of SAT traces illustrating the problem of identifying incomplete fragments. Small fragments are depicted in yellow, positive matches in red, negative but nearby matches in blue. In **C** and **D**, the small fragment is conclusively identified. In **F** and **G**, the correct SAT is assigned, but numerous other positive (false) matches are nearby, barely discriminated. In **E** and **H**, the fragment is very small and close to the cell bodies; the classification for BLVp1:1 is correct only because of the lack of other potentially positive matches. In **H**, only the enclosing lineage name (i.e., CM1), but not the specific SAT, is identified.

### Robustness: lineage identification in a glia-less mutant brain

The ablation of glial cells results in severely deformed brains, affecting the growth and pathfinding of SATs (Spindler et al., 2009). To test the robustness of our SAT identification algorithm, we traced and annotated 20, 25, 30, and 42 SATs in four glia-less third instar brain hemispheres (UAS-hid, UAS-rpr; Nirvana2-GAL4,UAS-GFP; tubGAL80[ts], confocal image stacks kindly provided by Shana Spindler, University of California, Los Angeles, CA), including the subset of lineages analyzed for fasciculation and growth defects in (Spindler et al., 2009) (Fig. 8). Despite SAT defects in the absence of neuropile glia, all SATs were classified correctly except in 1–4 lineages per brain, corresponding to lineages with a very high glia-association score (BAmas, DALcm1, DALcl1, and CP1). The BLD1–4 lineages, whose SAT is



**Figure 10.** Assigning primary neurons to SATs. Confocal z-projection (**A**) and confocal sections (**D**, **E**) of a larval brain labeled with antineurotactin (BP106; red) and anti-GFP (green), visualizing the set of dopaminergic neurons which expressed a UAS–GFP construct driven by tyrosine hydroxylase-specific promoter. In **A**, two clusters of dopaminergic neurons in the dorsomedial (p-DPM) and posterior (p-CP) brain; **D** shows an additional group (p-BLV) located further anteriorly and ventrally. All of these clusters represent primary neurons, as evidenced by the fact that their neurites have highly branched projections in the neuropile. As primary neurons, the dopaminergic cells project their axons adjacent or close to the SATs of secondary neurons, but not within the SATs themselves. Nevertheless, tracing and scoring the proximal unbranched segments of the primary neurons using NIT results in several positive matches. For the posterior neurons, the closest matches are DPLp2 and CP2/3:1 (**B**, **C**, and table). Each of the three individual BLV primary neurons (*p-BLV* in **D**) can be related by position and projection to one of the three BLVa SATs (**E**); the classifier suggests a collection of positive matches for BLVa SATs (data not shown). Only one cluster of dopaminergic primary neurons (labeled with a question mark) cannot be related to a SAT.

short and overlapping, and for which the classifier has a higher error rate in wild type (Table 2), were identified only at the lineage group level. The correct identification of a lineage like CP1 in three brains but not in the fourth may be explained by the incomplete penetrance of the heat-shock-induced glia-less phenotype.

In accordance with the findings of Spindler et al. (2009), we observe three types of pathfinding errors: fusion of the proximal segment of the SAT in sister lineages (the four mushroom body lineages, DPLal1–3, DAMv1/2, and BLD1–4) (Fig. 8C–E), shorter terminal projection (BAmas, DALcm1 and DALcl1) (Fig. 8F), and complete misrouting (CP1) (Fig. 8G) [see Spindler et al. (2009), their Table 1]. Only complete misrouting prevented the classifier from suggesting appropriate SAT annotations. Of note, the absence of glial cells severely disrupted the formation of the optic lobe, which lays immediately adjacent to the BLV, BLD, and BLA groups; yet, SATs of the BLV and BLA group were identified correctly.

The success in identifying nearly all SATs in a severely disrupted third instar brain suggests that the combination of a simple eight-landmark-based registration approach and the parameters used by NIT are robust enough for reliable SAT identification.

#### Flexibility: SAT identification with incomplete traces

Secondary axon tracts show a temporally and spatially dynamic expression of molecular markers. Some proteins, at a given stage, may be only found in part of the neuron. As a result, an SAT traced for identification may appear labeled over a much shorter length than that of its correspondent in the database. The example we present here is a larval brain labeled with an antibody against DE-cadherin (Fig. 9). This adhesion protein is transiently expressed in newly born secondary neurons and therefore visualizes SATs; it is also expressed on glial processes, which prevents one from following SATs in the neuropile.

On the other hand, SATs may be substantially longer than the traces by which they are represented in our database. This applies for the SATs of all lineages with a commissural projection. Because SATs of commissural lineages fasciculate with their contralateral counterparts, they cannot be traced beyond the midline in preparations that are globally labeled with antineurotactin. However, using clonal labeling techniques, these lineages may be visualized in their entirety.

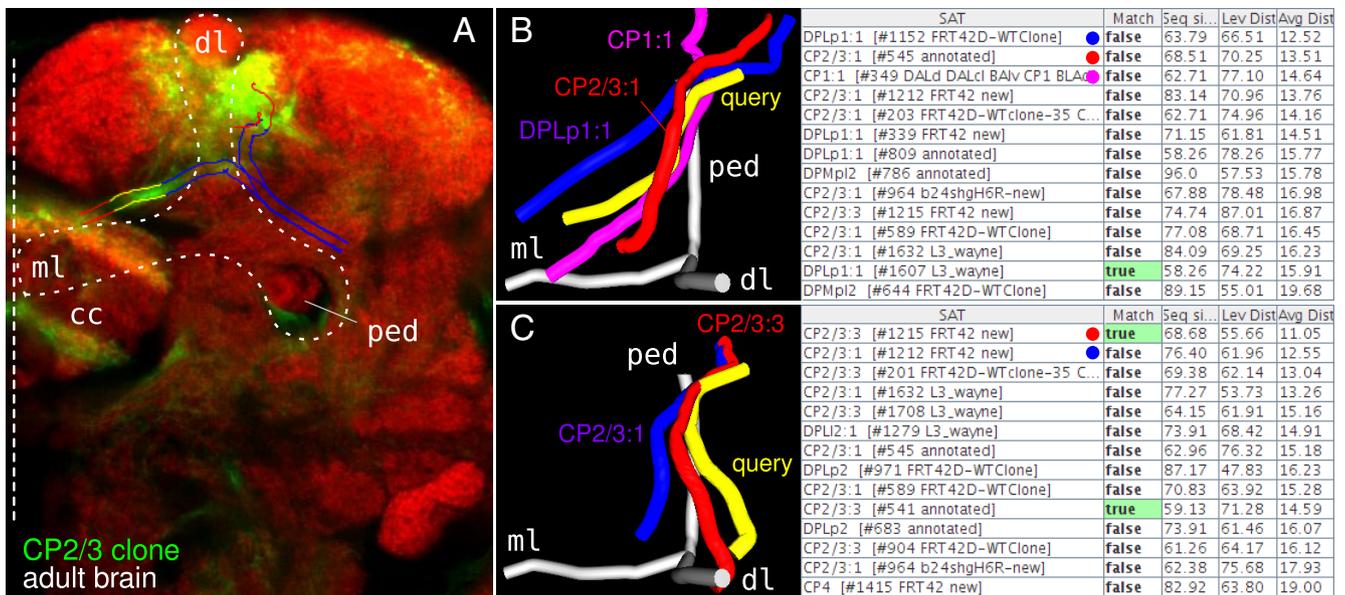
The NIT would deliver inappropriate results in most cases when scoring a SAT trace considerably shorter or longer than its true correspondent in the database. Realizing that the true SAT trace correspondence must lack large sequences of insertions or deletions, we devised a strategy consisting in performing global sequence alignment of the shortest of either the query or the reference SAT trace over all possible continuous longer trace subsequences of the shorter length (Fig. 9B).

To test the reliability of the assignment for shorter SAT traces, we manually annotated SATs in an anti-DE-cadherin-labeled brain. For the 11 lineages analyzed in the test brain (Fig. 9), nine were conclusively identified. For the remaining two, only the lineage name but not the specific SAT was identified (Fig. 9H). The random forest classifier presented an increased number of false positives. These arise as a result of very good matches with multiple SATs, because the short fragment may be common to multiple SATs, when the latter join common brain tracts, or near the cell bodies before diverging significantly (Fig. 9H).

For longer SAT traces, we used the flip-out clones of secondary lineages which aided in the construction of the database and which cross the midline or are considerably longer than the span of the neurotactin-labeled fraction of the SAT. All 12 longer SAT traces tested were correctly identified by the classifier, with 0–6 false positives for each.

#### Flexibility: identifying primary neurons and adult brain SATs with third instar SATs as reference

Primary neurons project a PAT into the neuropile (Nassif et al., 1998; Younossi-Hartenstein et al., 2006) where they arborize pro-



**Figure 11.** Identifying lineages in the adult brain using third instar annotated SATs. **A**, Confocal section of an adult *Drosophila* brain hemisphere, labeled with synaptotagmin (red). One lineage is labeled by the MARCM technique (in green); its two visible SATs were traced manually. Manual traces are indicated in blue (path above current slice), yellow (path within visible slice), and red (path below current slice). Dotted vertical line indicates midline; cc, central complex. **B, C**, Results of scoring the SATs of the adult brain against all SATs in the larval reference library. The tips and branch point of the mushroom body (dotted outline in **A**) provide four fiduciary points for registration of the traced SATs into the reference library coordinate space. The two tables show the top 14 results only, sorted by median Euclidean distance (Avg Dist column). Three-dimensional views show selected results from tables, indicated with corresponding colored dots. The Random Forest classifier labels the matches as true or false. The first GFP-labeled SAT (top table) results in zero true positives, one false positive, and five false negatives. The second GFP-labeled SAT (bottom table) results in two true positives, zero false positives, and three false negatives. In this case (**C**), the identity is clear: CP2/3:3. In the first case (**B**), without any true positives, the second and third results suggest CP2/3:1 over the first result, DPLp1:1. Visual inspection of the SAT traces in 3D offers further support for this decision. The possibility of annotating SATs in the adult brain using the third instar library of SATs as reference and only four fiduciary points suggests that the relative position of SATs lowest-order axonal branches do not change significantly from late larva to adult. dl indicates dorsal lobe; ml, medial lobe; ped, peduncle.

fusely, forming the larval neuropile. Secondary neurons, which develop throughout larval stages, form a SAT that follows the PAT into the neuropile (Larsen et al., 2009). We traced the proximal segment of the PAT for four individual primary neurons and two clusters of three primary neurons labeled with antidopamine in third instar brains. Then we ran NIT on the traces to identify SATs that follow similar trajectories (Fig. 10). We found that the four individual primary neurons were followed by four distinct SATs (BLVa1, a2 and a3; and DPMm1:4) (Fig. 10A–E), and one cluster of three primary neurons was followed by one SAT (CP2/3:1) (Fig. 10A–C). The remaining dorsomedial cluster of three primary neurons was not followed by any SAT. While primary neuron identification results were not and cannot be conclusive when comparing with SATs, NIT cuts down search to approximately three SATs in each case, suggesting between two and six positive matches with repetitions (Fig. 10B, C).

Secondary neurons arborize during pupal stages, forming the largest fraction of the adult fly brain. The low-order branch of the SAT remains recognizable for at least a subset of lineages with antineuroglial antibody or with genetically targeted GFP-labeling. We traced the two SATs of one of the two CP2/3 sister lineages (Fig. 11A) and ran NIT against the third instar SAT reference library. There were two true positives for one SAT (CP2/3:3) (Fig. 11C). While the classifier did not find any positives for the other SAT (CP2/3:1) (Fig. 11B), the top five results contained three instances of CP2/3:1. The 3D registration with only four fiduciary landmarks (the four mushroom body corners) proved sufficient for NIT to provide good annotation suggestions for the adult CP2/3 lineage SATs.

#### Effect of 3D registration on SAT identification

The algorithmic classification of an anatomical structure depends both on the recognition and quantification of its geometrical properties and on the correct elucidation of its relative spatial location to other potentially similar structures. We have approached the spatial positioning problem with a 3D registration strategy based on eight fiduciary points, which are easily recognizable and evenly distributed (Fig. 3). Each brain was mounted in a different orientation, and thus artifactual deformations were introduced in a nonsystematic way. However, SAT identification across brains is accurate in wild-type brains, in the four mutant glia-less brains, and in the D-cadherin labeled brain (Tables 1–4).

We have observed the effect of insufficient brain-to-brain registration for some SATs as an increased distance between correspondent SATs across the five reference brains (e.g., BALv) (Fig. 5C). The fact that no other lineage presents such large distances between its cognates (Fig. 5) suggests that volume registration with only eight landmarks is sufficient for the identification of protocerebral and deutero-cerebral lineages. All tritocerebral lineages, which lay beyond the volume circumscribing the eight fiducial points, cannot be reliably identified.

The better the 3D brain registration, the more reliable the recognition of any brain structure will be. Our approach is flexible regarding the number of fiduciary points. For many lineages, using only the four corner points of the mushroom body is sufficient for accurate recognition. The additional four fiduciary points greatly enhance the fidelity for many other lineages, particularly the basal and lateral groups. If needed, further fiduciary points could be added. The strength of our approach is that it

performs reliably even if “perfect” registration is not possible (e.g., across mutant brains) (Fig. 8).

## Discussion

We presented the NIT for the quantitative measurement of similarity between a pair of neurite traces in the same reference space. We built a reference library of traced and annotated SATs labeled with antineurotactin at the third instar developmental stage of *Drosophila* larva, which is publicly available at <http://t2.ini.uzh.ch/nit/>. We applied the tool for the semiautomatic annotation of SATs in wild-type and extended and refined the secondary lineage nomenclature by Peraanu and Hartenstein (2006). We quantified the accuracy of the automatically suggested annotations, and we found it to be between 93% and 99%. We tested the automatic annotation in mutant third instar brains and in brains labeled with markers other than antineurotactin and found the accuracy to be very high. Remarkably, we were able to use NIT for the discovery of primary neurons associated with SATs and for the identification of SATs in the adult brain. Upon recognition of a few fiducial points for 3D registration, NIT significantly decreases the time it takes for identifying a neuron or a lineage. Therefore, our tool may provide the means to link digital atlases of neurons and neuronal lineages across developmental stages of *Drosophila*, from late embryo to the adult.

### The accuracy of SAT identification

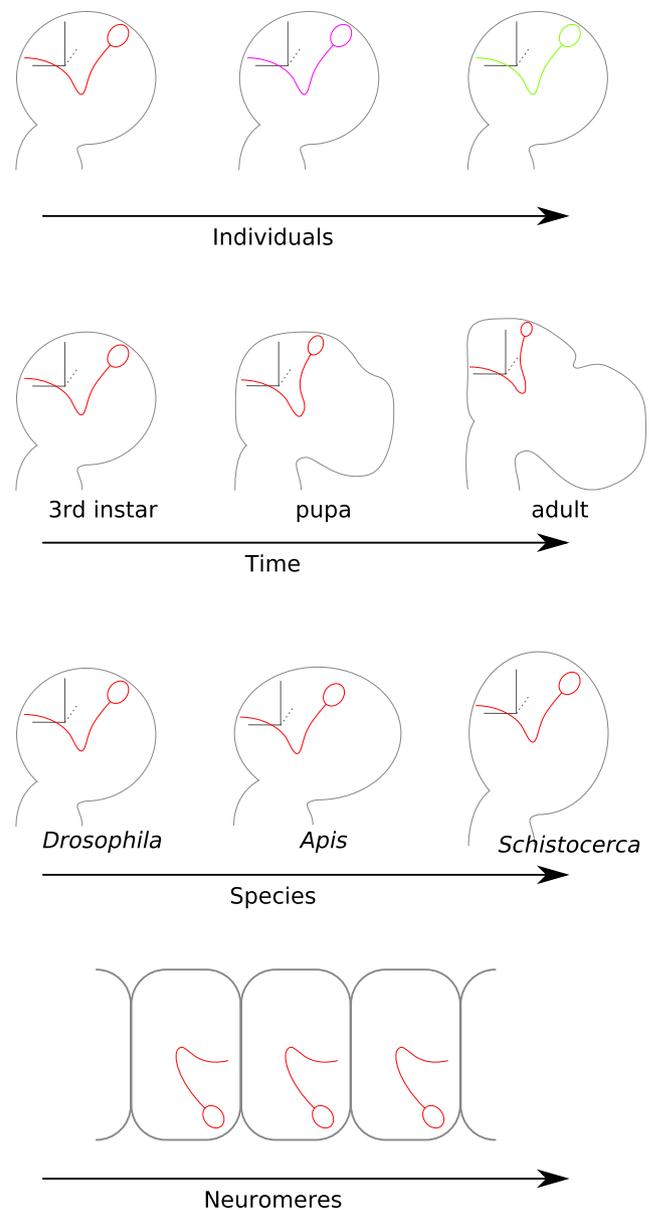
There exist differences in the accuracy with which individual lineages could be identified with their proper cognate in the reference database. Shorter SATs are likelier to share their trajectory with other SATs and are less reliably identified. Lineages BLD1-4, BLAd2-3, and BLA1 present relatively short SATs that project into the compact transverse superior fascicle. Beyond small differences in their relative position, their distinctive feature is the position of the cell body cluster. Consequently, SATs in the lineage groups BLD and BLA are among the hardest to annotate by hand and among the least well discriminated by the median Euclidean distance parameter (Table 2).

The second type of errors in SAT identification occurs in cases where multiple lineages share, for substantial parts of their trajectories, a compact fascicle, before diverging from each other to form more discrete terminal segments. This happens for the CM lineage group, all of which share the longitudinal central or longitudinal superior medial fascicle. SATs of the CM groups are also hard to annotate by hand in neurotactin-labeled brains, given (1) the elevated number of SATs per lineage, and (2) the high density of SATs from multiple lineages near the future location of the central complex. The generation of GFP-labeled flip-out clones was essential for the manual annotation of CM lineages.

The sister lineages represent a third set of instances with problematic manual and algorithmic identification. The prime examples are DAMd2 and DAMd3, which the trained classifier cannot discriminate from each other, presenting results with numerous false positives. These lineages are particularly hard to annotate by hand, and their annotation may remain ambiguous until sufficient GFP-labeled flip-out clones indicate individual characteristics.

### Stereotypy and SATs

The concept of morphological stereotypy may be defined as the degree of variability of a structure across individuals as measured under a given set of conditions. Our SAT semiautomatic annotation approach was made possible by the observed strong stereo-



**Figure 12.** Dimensions of comparisons. Schematic diagram of SAT comparisons in four different dimensions: individuals at the same developmental time; individuals across developmental times; across different insect species; and across serially homologous parts of the nervous system such as ventral nerve cord abdominal segments. Diagrams in the first three rows schematize a right brain hemisphere with a colored SAT, and the mushroom body in black as a reference coordinate system.

typy of neuronal components in *Drosophila* and insect brains in general. In particular, neuroblasts (Technau, 2008) and neuronal cell bodies (Hiesinger et al., 2006), primary neuronal lineages (Nassif et al., 1998), secondary neuronal lineages (Ito and Hotta, 1992; Ito et al., 1997; Peraanu and Hartenstein, 2006), olfactory neurons (Jefferis et al., 2007), and neuropile compartments (Younossi-Hartenstein et al., 2003; Jenett et al., 2006) have been described as highly stereotypical in their position, dimensions, and numbers.

Stereotypy, as defined, is a function of measurement conditions. Measuring stereotypy to generate consensus SAT traces, or to identify the same SAT in different brains, requires nonlinear 3D registration, which eliminates global and local differences in brain size and shape. These differences originated in artifactual

deformations induced by sample preparation, in the exact developmental time at which each larva was fixed and dissected, and in phenotypical differences among individuals. Under the condition of equal coordinate space, corresponding SATs from different wild-type individuals are very similar yet not identical. We have observed a nonhomogeneous distribution of the residual variability (Fig. 6), defined as the remaining variability between two structures after bringing them into approximately the same coordinate space.

The proximal segment of a SAT, located close to the cell bodies, is generally more variable than the rest, in agreement with reports on relatively high variability in the location of the neuroblasts and neuronal cell bodies (Pereanu and Hartenstein, 2006). Neuroblasts and their attached cell body clusters can vary up to approximately the diameter of a lineage ( $\sim 10\text{--}12\ \mu\text{m}$ ). By contrast, we find a much lower variability of the middle and distal segments of the SAT, which indicates a stricter regulation of axon pathfinding and positioning within the neuropile (Fig. 6).

### Identification and quantification of mutant secondary lineages

We report a high accuracy in the recognition of SATs at third instar larva. We tested the robustness of our method by attempting the identification of SATs in mutant severely deformed glia-less brains. Our method correctly identified numerous SATs, indicating a preservation of relative positions of SATs within the neuropile despite lacking glial cells. The difficulty of manually annotating SATs in mutants is reduced to determining the position of fiduciary points, then trace and compare a SAT to all SATs in the annotated library. The systematic application of our method to the SATs of glia-less brains highlighted some SATs without conclusive identification, suggesting these lineages were strongly affected, in accordance with their close association with glial sheets (Spindler et al., 2009). The sequence analysis approach used by our method indicates the presence of tract segment deletions (such as the missing terminal segment in BAmas1/2) (Fig. 8G) or complete misrouting (Fig. 8F).

This robustness of NIT will facilitate studies of mutant phenotypes. Current methods for quantifying phenotypic changes in the brain were mostly qualitative and restricted to parts where changes are most obvious, like the mushroom body (Heisenberg et al., 1985; Heisenberg, 1998). Quantitative mutant analysis can now be extended to all central brain lineages.

### Bootstrapping digital atlases of different developmental stages, species, and imaging modalities

How far can 3D registration be pushed to obtain valuable suggestions on the identity of a SAT? Beyond individuals, our approach enables identification of primary and secondary lineages across developmental time, nerve cord segments, and species (Fig. 12). Easily identifiable lineages, such as those forming the mushroom body, antennal lobe, or central complex, have been identified across all insect taxa (Williams and Boyan, 2008; Strausfeld et al., 2009). With appropriate fiduciary points, NIT could relate lineages from different species to those of *Drosophila*.

Efforts are underway to use the same techniques by which lineages were reconstructed in the larva to follow neuronal lineage differentiation throughout pupal stages into the adult. Our exploratory data (Fig. 11) indicate that the brain, despite undergoing massive growth after arborization of secondary lineages, does not change substantially in the relative position of internal components. However, certain changes occur, being nontrivial to identify adult lineages using their larval instances. For example,

certain SATs grow massively; more generally, the clustering of somata changes as the brain cortex expands and simultaneously becomes thinner (Larsen et al., 2009). As a result, proximal SAT segments change in direction. For best accuracy, and to detect newly developed or lost SATs in the pupal period, we envision multiple digital atlases of SATs for several pupal stages and the adult brain. Our quantitative analysis of similarities between SATs will be invaluable in establishing lineage identity through time.

### The use of NIT for the semiautomated mapping of neurons

Systematic approaches to obtain markers for eventually every cell of the brain and ventral nerve cord have been initiated for the *Drosophila* brain (Pfeiffer et al., 2008). Such markers are crucial for the usage of *Drosophila* as a model for neural function, development, or pathology.

Neurons of the vertebrate brain are defined topologically in relation to rich spatial frameworks of reference, composed of compartments (e.g., “lateral geniculate nucleus”) and tracts (e.g., “olivocerebellar tract”). Neurons in invertebrate brains have been classified based on cell body position (“dorsomedial group”) or special attributes (“giant fiber neuron”; “optic lobe pioneers”), lacking detailed reference frameworks. This non-systematic classification is insufficient for the comparison of different sets of neurons. We propose neuronal lineages as a high-resolution topological framework for single neuron identification.

With NIT, traced low-order branch segments of labeled neurons may be used for the assignment of neurons to their enclosing lineages with high reliability, providing them with a genetic address. The lineage represents an envelope; the knowledge we have about the envelope will overlap to a large extent with the individual neuron enclosed in it.

### References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Ascoli GA, Donohue DE, Halavi M (2007) NeuroMorpho.Org: a central resource for neuronal morphologies. *J Neurosci* 27:9247–9251.
- Ascoli GA, Brown KM, Calixto E, Card JP, Galván EJ, Perez-Rosello T, Barrionuevo G (2009) Identified CA3b interneurons reveals robust local geometry and distinct cell classes. *J Comp Neurol* 515:677–695.
- Bello B, Izergina N, Caussinus E, Reichert H (2008) Amplification of neural stem cell proliferation by intermediate progenitor cells in *Drosophila* brain development. *Neural Dev* 3:5.
- Bossing T, Udolph G, Doe CQ, Technau GM (1996) The embryonic central nervous system lineages of *Drosophila melanogaster*. I. Neuroblast lineages derived from the ventral half of the neuroectoderm. *Dev Biol* 179:41–64.
- Brandt R, Rohlfing T, Rybak J, Kroficzek S, Maye A, Westerhoff M, Hege H, Menzel R (2005) Three-dimensional average-shape atlas of the honeybee brain and its applications. *J Comp Neurol* 492:1–19.
- Breiman L (2001) Random Forests. *Machine Learn* 45:5–32.
- Condron BG (2008) A freeware java tool for spatial point analysis of neuronal structures. *Neuroinformatics* 6:57–61.
- Cornbrooks C, Bland C, Williams DW, Truman JW, Rand MD (2007) Delta expression in post-mitotic neurons identifies distinct subsets of adult-specific lineages in *Drosophila*. *Dev Neurobiol* 67:23–38.
- Dubes R, Jain A (1988) Algorithms for clustering data. Upper Saddle River, NJ: Prentice Hall.
- Dumstrei K, Wang F, Nassif C, Hartenstein V (2003) Early development of the *Drosophila* brain: V. Pattern of postembryonic neuronal lineages expressing DE-cadherin. *J Comp Neurol* 455:451–462.
- Felsenstein J (1989–2009) PHYLIP (Phylogeny Inference Package) version 3.69.
- Fung S, Wang F, Spindler SR, Hartenstein V (2009) *Drosophila* E-cadherin and its binding partner Armadillo/b-catenin are required for axonal pathway choices in the developing larval brain. *Dev Biol* 332:371–382.

- Geinisman Y, Gundersen HJ, van der Zee E, West MJ (1996) Unbiased stereological estimation of the total number of synapses in a brain region. *J Neurocytol* 25:805–819.
- Grover D, Tavaré S (2009) Finding behavioral motifs in fly trajectories. *Comm Inform Syst* 9:283–294.
- Hartenstein V, Campos Ortega J (1997) The embryonic development of *Drosophila melanogaster*, Ed 2. Berlin: Springer.
- Heisenberg M (1998) What do the mushroom bodies do for the insect brain? An introduction. *Learn Mem* 5:1–10.
- Heisenberg M, Borst A, Wagner S, Byers D (1985) *Drosophila* mushroom body mutants are deficient in olfactory learning. *J Neurogenet* 2:1–30.
- Heumann H, Wittum G (2009) The tree-edit-distance, a measure for quantifying neuronal morphology. *Neuroinformatics* 7:179–190.
- Hiesinger PR, Zhai RG, Zhou Y, Koh TW, Mehta SQ, Schulze KL, Cao Y, Verstrecken P, Clandinin TR, Fischbach KF, Meinertzhagen IA, Bellen HJ (2006) Activity-independent prespecification of synaptic partners in the visual map of *Drosophila*. *Curr Biol* 16:1835–1843.
- Ito K, Hotta Y (1992) Proliferation pattern of postembryonic neuroblasts in the brain of *Drosophila melanogaster*. *Dev Biol* 149:134–148.
- Ito K, Awano W, Suzuki K, Hiromi Y, Yamamoto D (1997) The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development* 124:761–771.
- Jefferis GS, Marin EC, Stocker RF, Luo L (2001) Target neuron prespecification in the olfactory map of *Drosophila*. *Nature* 414:204–208.
- Jefferis GS, Potter CJ, Chan AM, Marin EC, Rohlfsing T, Maurer CR Jr, Luo L (2007) Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* 128:1187–1203.
- Jenett A, Schindelin JE, Heisenberg M (2006) The Virtual Insect Brain protocol: creating and comparing standardized neuroanatomy. *BMC Bioinformatics* 7:544.
- Jiang X, Bunke H, Abegglen K, and Kandel A (2002) Curve morphing by weighted means of strings. Paper presented at the Sixteenth International Conference on Pattern Recognition (ICPR 2002), August, Quebec, Canada.
- Jurrus E, Hardy M, Tasdizen T, Fletcher P, Koshevoy P, Chien C, Denk W, Whitaker R (2009) Axon tracking in serial block-face scanning electron microscopy. *Med Image Anal* 1:180–188.
- Kumar A, Fung S, Lichtneckert R, Reichert H, Hartenstein V (2009) The arborization pattern of engrailed-positive neural lineages reveal neuro-mere boundaries in the *Drosophila* brain neuropile. *J Comp Neurol* 517:87–104.
- Larsen C, Shy D, Spindler SR, Fung S, Pereanu W, Younossi-Hartenstein A, Hartenstein V (2009) Patterns of growth, axonal extension and axonal arborization of neuronal lineages in the developing *Drosophila* brain. *Dev Biol* 335:289–304.
- Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.
- Macke JH, Maack N, Gupta R, Denk W, Schölkopf B, Borst A (2008) Contour-propagation algorithms for semi-automated reconstruction of neural processes. *J Neurosci Methods* 167:349–357.
- Mishchenko Y (2009) Automation of 3D reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs. *J Neurosci Methods* 176:276–289.
- Nassif C, Noveen A, Hartenstein V (1998) Embryonic development of the *Drosophila* brain. I. Pattern of pioneer tracts. *J Comp Neurol* 402:10–31.
- Nassif C, Noveen A, Hartenstein V (2003) Early development of the *Drosophila* brain: III. The pattern of neuropile founder tracts during the larval period. *J Comp Neurol* 455:417–434.
- Pereanu W, Hartenstein V (2006) Neural lineages of the *Drosophila* brain: a three-dimensional pattern of lineage location and projection at the late larval stage. *J Neurosci* 26:5543–5553.
- Pfeiffer BD, Jenett A, Hammonds AS, Ngo TT, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Laverty TR, Mungall C, Svirskas R, Kadonaga JT, Doe CQ, Eisen MB, Celniker SE, Rubin GM (2008) Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc Natl Acad Sci U S A* 105:9715–9720.
- Schaefer S, McPhail T, Warren J (2006) Image deformation using moving least squares. *ACM Trans Graphics* 25:533–540.
- Schmidt H, Rickert C, Bossing T, Vef O, Urban J, Technau GM (1997) The embryonic central nervous system lineages of *Drosophila melanogaster*. II. Neuroblast lineages derived from the dorsal part of the neuroectoderm. *Dev Biol* 189:186–204.
- Schmitt S, Evers J, Duch C, Scholz M, Obermayer K (2004) New methods for the computer-assisted 3-D reconstruction of neurons from confocal image stacks. *Neuroimage* 4:1283–1298.
- Sholl D (1953) Dendritic organization in the neurons of the visual and motor cortices of the cat. *J Anat* 87:387–406.
- Smith TF, Waterman MS (1981) Identification of common molecular subsequences. *J Mol Biol* 147:195–197.
- Spindler SR, Ortiz I, Fung S, Takashima S, Hartenstein V (2009) *Drosophila* cortex and neuropile glia influence secondary axon tract growth, path-finding, and fasciculation in the developing larval brain. *Dev Biol* 334:355–368.
- Strausfeld NJ, Sinakevitch I, Brown SM, Farris SM (2009) Ground plan of the insect mushroom body: functional and evolutionary implications. *J Comp Neurol* 513:265–291.
- Technau G (2008) Brain development in *drosophila melanogaster*. Austin, TX: Landes Bioscience.
- Urbach R, Technau GM (2004) Neuroblast formation and patterning during early brain development in *Drosophila*. *BioEssays* 26:739–751.
- Van Pelt J, Uylings HB, Verwer RW, Pentney RJ, Woldenberg MJ (1992) Tree asymmetry—A sensitive and practical measure for binary topological trees. *Bull Math Biol* 54:759–84.
- Viana MP, Tanck E, Beletti ME, Costa Lda F (2009) Modularity and robustness of bone networks. *Mol Biosyst* 5:255–261.
- Wagner R, Fischer M (1974) The string-to-string correction problem. *J Assoc Comp Machin* 21:168–173.
- Weaver CM, Hof PR, Wearne SL, Lindquist WB (2004) Automated algorithms for multiscale morphometry of neuronal dendrites. *Neural Comput* 16:1353–1383.
- Williams JL, Boyan GS (2008) Building the central complex of the grasshopper *Schistocerca gregaria*: axons pioneering the w–z tracts project onto the primary commissural fascicle of the brain. *Arthropod Struct Dev* 37:129–140.
- Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques. San Francisco: Morgan Kaufmann.
- Younossi-Hartenstein A, Salvaterra PM, Hartenstein V (2003) Early development of the *Drosophila* brain IV. Larval neuropile compartments defined by glial septa. *J Comp Neurol* 455:435–450.
- Younossi-Hartenstein A, Nguyen B, Shy D, Hartenstein V (2006) Embryonic origin of the *Drosophila* brain neuropile. *J Comp Neurol* 497:981–998.
- Yu HH, Chen CH, Shi L, Huang Y, Lee T (2009) Twin-spot MARCM to reveal the developmental origin and identity of neurons. *Nat Neurosci* 12:947–953.