Identification of Neuronal Lineages in the *Drosophila* Peripheral Nervous System with a "Digital" Multi-spectral Lineage Tracing System

Macy W. Veling[1,2,7], **Ye Li**[3,7], **Mike T. Veling**[4,5,7], **Christopher Litts**[1], **Nigel Michki**[6], **Hao Liu**[1,3], **Bing Ye**[1,2,3,8,*], **Dawen Cai**[3,6,*]

[1]Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109, USA

[2]Cellular and Molecular Biology Graduate Program, University of Michigan, Ann Arbor, MI 48109, USA

[3]Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

[4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

[5]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

[6]Department of Biophysics, University of Michigan, Ann Arbor, MI 48109, USA

[7]These authors contributed equally

[8]Lead Contact

SUMMARY

Elucidating cell lineages provides crucial understanding of development. Recently developed sequencing-based techniques enhance the scale of lineage tracing but eliminate the spatial information offered by conventional approaches. Multi-spectral labeling techniques, such as Brainbow, have the potential to identify lineage-related cells *in situ*. Here, we report nuclear Bitbow (nBitbow), a "digital" version of Brainbow that greatly expands the color diversity for scoring cells, and a suite of statistical methods for quantifying the lineage relationship of any two cells. Applying these tools to the *Drosophila* peripheral nervous system (PNS), we determined lineage relationship between all neuronal pairs. This study demonstrates nBitbow as an efficient tool for *in situ* lineage mapping, and the complete lineage relationship among larval PNS neurons opens new possibilities for studying how neurons gain specific features and circuit connectivity.

In Brief

Veling et al. report a multi-color labeling system and statistical methods for mapping cell lineages. They identify the lineage relationship of all neurons in the peripheral nervous system of *Drosophila* larvae and show the utility of this technique in mapping neurons in the CNS.

Graphical Abstract
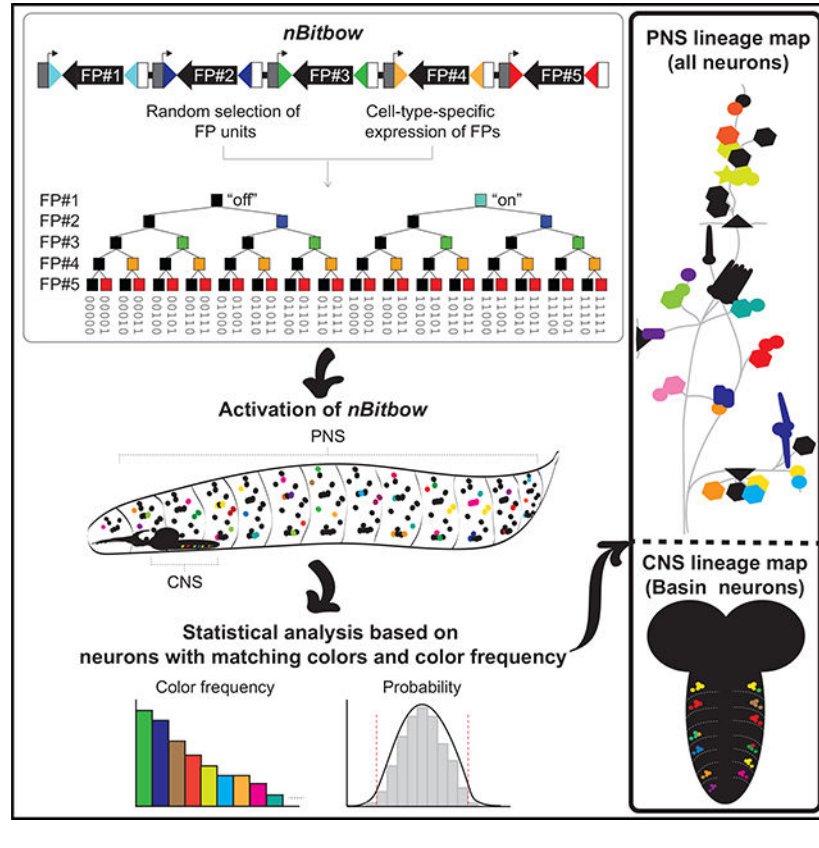
INTRODUCTION

Cell lineage, which denotes the developmental history of a cell, provides the conceptual framework for understanding organism formation (Papaioannou, 2016; Stent, 1985). For example, identifying the lineage relationship among neurons is essential for understanding how neurons gain specific physiological, morphological, and neurochemical features and proper circuit connectivity (Hobert and Kratsios, 2019; Lacin et al., 2019; Lee, 2017). Modern molecular genetic techniques have led to evolutionary improvement in lineage tracing from classic techniques involving dye filling or cell transplantation (Woodworth et al., 2017). For instance, sequencing-based methods can distinguish hundreds to thousands of uniquely barcoded lineages (Raj et al., 2018; Schmidt et al., 2017; Spanjaard et al., 2018). However, sequencing-based methods do not offer *in situ* spatial relationships of these lineages, because they require tissue disassembly. Moreover, the birth timing of cells within a lineage is difficult to resolve. In contrast, imaging-based methods, such as Brainbow, can preserve the spatial information *in situ* and permit live imaging, but their efficiencies are still

limited by labeling diversities and lack of statistical tools for unambiguous lineage tracing (Boulina et al., 2013; Cachero and Jefferis, 2011; Cai et al., 2013; Hadjieconomou et al., 2011; Hampel et al., 2011; Livet et al., 2007; Pan et al., 2013). Hence, there is an urgent need to create novel labeling, detection, and quantification methods that allow highly efficient lineage tracing while preserving spatial information.

Brainbow, a multi-spectral labeling technology, is designed to randomly express one of three or four fluorescent proteins (FPs) from a single cassette, thus creating stochastic labeling colors in neighboring cells or cell lineages (Boulina et al., 2013; Cachero and Jefferis, 2011; Cai et al., 2013; Hadjieconomou et al., 2011; Hampel et al., 2011; Livet et al., 2007; Pan et al., 2013). When more color variants are needed to uniquely label many cell lineages, more than one Brainbow cassette can be used to create differential expression levels of FPs. However, using color shades for lineage tracing is not always reliable. Distinguishing two color variants differing by subtle FP expression levels (e.g., color A generated from 2 YFP + 1 RFP + 1 CFP compares to color B generated from 1 YFP + 2 RFP + 1 CFP) can be challenging because of imaging noise (Cai et al., 2013). When applying Brainbow to trace cell lineages, daughter cells in the same lineage are assumed to inherit the same color generated in the mother stem cell. However, protein synthesis levels in daughter cells can be quite different. A more robust color generation mechanism would address these concerns and provide more reliable lineage tracing.

One way to generate more robust Brainbow lineage labels is to localize the same FPs to different subcellular compartments. Cytoplasmic membrane-targeted and nucleus-targeted FPs, expressed through genome integration by electroporated transposase, have been used to differentiate neighboring neuronal lineages in chick and mouse embryonic brains (García-Moreno et al., 2014; Loulier et al., 2014). However, transposase integrates varying numbers of targeting cassettes in different cells, making it difficult to estimate the probability of each label combination for quantitative analysis. Generating transgenic animals with a fixed number of labeling cassettes would solve this problem. For instance, the Raeppli strategy used 4 FPs to create up to $4 \times 4 = 16$ membrane and nucleus color combinations in transgenic *Drosophila* (Kanca et al., 2014). Another recombination mechanism, implemented in the CLoNe and the MultiColor FlpOut (*MCFO*) *Drosophila* lines, generates random colors by stochastic removal of the expression stops from each FP module (García-Moreno et al., 2014; Nern et al., 2015). For instance, a *MCFO* fly integrates 3 different stop-spaghetti monster GFPs (smGFPs) modules into 3 distinct genomic loci and generates up to $2^3 - 1 = 7$ smGFP combinations. However, expanding the color pool requires inserting more FP modules to additional genomic loci to prevent inter-module recombination. In addition, the optimal color outcome can be difficult to obtain from *CLoNe* and *MCFO* animals. This is because while low FLP activity results in simple single-marker colors, high FLP activity often results in expressing all FPs in most cells (Nern et al., 2015). In conclusion, the small unique color pools generated by the previously mentioned methods result in frequent observations that neighboring cells or lineages are labeled with the same color.

To overcome these limitations, we introduced a "digital" format of Brainbow for robust lineage tracing, which we call nuclear Bitbow (nBitbow). A single nBitbow cassette composites binary switches to independently determine an ON/OFF state of 5 FPs to

generate up to $2^5 - 1 = 31$ color variants. To facilitate the detection of the FPs independent of their intensities, the FPs are tagged for nuclear localization. We applied nBitbow to trace neuronal lineages in the *Drosophila* peripheral nervous system (PNS). We chose the fly PNS as a model because (1) individual PNS neurons are identifiable by their soma locations and neurite patterns; (2) PNS neuronal lineages reported in previous studies can serve as references for result comparison (Brewster and Bodmer, 1995, 1996); and (3) previous neuronal lineage mapping in PNS was incomplete because of technical challenges. Because there is a lack of statistical methods for color-based lineage tracing, we developed a lineage-relatedness test to determine the likelihood of two cells being lineage related based on their colors. Using this statistic test, we confirmed and rejected previously determined lineages, as well as revealing previously unidentified lineages. Moreover, applying nBitbow labeling at different developmental time points, we revealed the proximate birth timing of many PNS neuronal lineages. Furthermore, we demonstrated the feasibility of using nBitbow for lineage mapping in other systems by applying it to determine the lineage relationship of a subset of interneurons in the CNS.

RESULTS

nBitbow Design and Characterization

As opposed to previous Brainbow designs that expand color variations by recombining multiple cassettes to mix different intensity levels in each spectral channel, Bitbow, a digital form of Brainbow, use several (N) FPs and randomly assign each to an ON or OFF expression state upon FLP-mediated recombination (Figure 1A). As a result, Bitbow generates N-bit ($2^N - 1$, minus the all-OFF state) colors from a single cassette. We further optimized Bitbow for lineage tracing and generated a nuclear Bitbow (nBitbow) cassette containing 5 FPs (mAmetrine, mTFP, mNeonGreen, mKusabira-Orange2, and tdKatushka2), each of which was fused to a human histone H2B (hH2B) for nuclear tagging. The cDNA of each nuclear FP was positioned in the inverse direction and flanked by a pair of incompatible FLP recognition target (FRT) sites, thus permitting spinning the FPs in forward or inverse orientation upon FLP recombination (Cai et al., 2013; McLeod et al., 1986; Schlake and Bode, 1994; Turan et al., 2010; Volkert and Broach, 1986). Unlike previous Brainbow designs, recombination of each nuclear FP in nBitbow was independent of each other, resulting in up to $2^5 - 1 = 31$ unique colors by a single cassette (Figure 1B). Finally, an upstream activating sequence (UAS) and a p10 baculovirus poly-adenylation sequence were placed upstream and downstream, respectively, to each of the five FP recombination units to allow strong FP expression when driven by a Gal4 driver (Pfeiffer et al., 2012). We chose FPs whose native signals are bright and stable so that immunostaining is not required. In addition, spectral imaging and linear unmixing were applied and optimized for separating the five fluorescent signals to avoid signal bleed-through (Figure 1C).

To evaluate the labeling efficiency of nBitbow in the *Drosophila* larval PNS, we first combined a transgene encoding heat shock-inducible flippase (*hsFLP*) and the pan-neuronal driver elav-Gal4 to determine how often neurons expressed colors in the third-instar larvae. Without heat shock (no hs), an average of $2.89\% \pm 0.90\%$ (mean $\pm$ SEM) of PNS neurons showed colors when the larvae were raised at 23C (Figure 2A), likely due to leaky FLP

expression, because we did not observe FP expression in the absence of the FLP transgene (no FLP) (Golic and Lindquist, 1989). Lowering the developmental temperature to 18C did not reduce this leaky FLP activity (data not shown). Induction of FLP expression by a 30-min heat shock at 37C that started at 3, 4, or 5 h after egg laying (AEL) led to significant increases of the number of PNS neurons that expressed colors (up to 18.62% ± 2.74%) (Figure 2A). Heat shock that started at 2 h AEL turned on color expression in a similar number of neurons compared with the non-heat shock condition (Figure 2A), possibly because of either low levels of FLP expression or poor accessibility of FLP to the nBitbow construct at very early stages of embryogenesis. Interestingly, we observed that the efficiency of turning on colors at 6 h AEL decreased to the level of non-heat shock condition. This may indicate a lower efficiency of FLP recombination in the post-mitotic cells (detailed later).

Because different FRT sites exhibit different efficiencies of FLP-induced recombination (Schlake and Bode, 1994; Turan et al., 2010), we quantified the expression frequency of individual FPs. mNeonGreen displayed the highest expression frequency, followed by mTFP, mKusabira-Orange2, tdKatushka2, and mAmetrine (Figure 2B). Next, we quantified the frequency of neurons expressing one, two, three, four, or all five FPs. We found that such frequency followed a Poisson distribution from expressing one FP to five FPs. Chi-square goodness-of-fit test confirmed that such distribution (actual) had no significant difference from the theoretical distribution (expected) (Figure 2C). Thus, the expression of one FP does not appear to increase the chance of expressing other FPs, suggesting that the recombination of distinct FRT sites is independent of each other. Finally, the expression frequencies of all FP combinations (i.e., color variants) are quantified for our lineage-relatedness test (Figure 2D) (detailed later). Despite that different combinations exhibited different frequencies, we observed 27 of the 31 expected colors after heat shock-induced FLP expression.

One potential drawback of nBitbow for lineage tracing is that the spinning recombination design may cause color changes over time if the leaky FLP activity is persistent in a cell. As a result, cells with altered colors in the same lineage may be mis-identified as being in distinct lineages. To test this, we applied the same heat shock protocol and examined nBitbow color labeling in the CNS, because neuroblasts (NBs) in the CNS divide many more times than the sensory organ precursors (SOPs) in the PNS to generate cell clusters with clear boundaries (Birkholz et al., 2015; Truman and Bate, 1988). We found that most cells in each cluster were labeled in the same nBitbow color (Figure 2E). We occasionally observed a few scattered cells labeled by different colors in a cluster and rarely observed a cell cluster that was labeled by two distinctly colored clones (Figure 2F). This result is surprising, because every neuroblast division generates a ganglion mother cell (GMC) that produces 2 daughter cells. As a consequence, similar numbers of neuroblast clones and 1- or 2-cell clones are expected. We observed a very small number of 1-cell or 2-cell clones, which indicates that for unknown reasons, the nBitbow labeling induced by embryonic heat shock mostly happens in neural stem/progenitor cells and that color changes caused by persistent FLP activity are minimal. We also observed that nBitbow had lower labeling efficiency in the PNS when heat-shocked at 6 h AEL (Figure 2A), a stage that has more post-mitotic cells. Altogether, these results suggest that nBitbow provides more unique

colors from a single cassette than previous variants of Brainbow and is reliable for lineage tracing in the *Drosophila* nervous system.

Identification of Neuronal Lineages in the *Drosophila* Larval PNS by nBitbow

Drosophila PNS neurons are generated through lineages that repeat in the abdominal hemisegments (Figure 3A) (Hartenstein and Campos-Ortega, 1984; Hartenstein et al., 1987). The SOPs in each hemisegment are born and delaminated from ectoderm around 3.5–7.5 h AEL to eventually give rise to 45 PNS neurons (Bodmer et al., 1989; Ghysen and O'Kane, 1989; Younossi-Hartenstein and Hartenstein, 1997). We therefore induced nBitbow recombination in PNS neurons by 30-min heat shock at 3–4 or 5–6 h AEL and recorded their colors in the third-instar larvae (Figure 3B). We expect to observe a significantly higher probability of two neurons being labeled in the same colors if they belong to the same lineage (i.e., inheriting the same colors from their SOP) (Figure 3C) than if they belong to different lineages (i.e., acquiring the same colors by chance).

The different expression frequencies of the color variants call for statistical methods that weigh these variants appropriately in their strengths for deducing the lineage relationship between two neurons (Figure 2D). Based on Cohen's kappa (k) statistics (Cohen, 1960; Dwass, 1957; Nichols and Holmes, 2002, 2007), which calculates the above-background match rate in non-uniform distributions like ours, we designed a randomized lineage-relatedness test—with the frequencies of each nBitbow color weighed in—to determine whether two neurons in a pair are lineage related. Ranging from −1 to 1, $\kappa = 1$ indicates the two neurons are always labeled by the same colors, $\kappa = 0$ indicates the neurons are labeled with the same colors by chance, and $\kappa = -1$ indicates the two neurons are always labeled with different colors. We calculated the κ value of every pair of PNS neurons in *Drosophila* larvae and determined whether each κ value is significantly different from the probability of labeling the same neuron pair with the same colors by chance in a randomization test (Figure 3D, and see STAR Methods for further details). Finally, a 2% false discovery rate (FDR) based on the Benjamini-Hochberg method was applied to confirm any pair of neurons that are lineage related (Benjamini and Hochberg, 1995).

Because some neurons are labeled at the third-instar larval stage without heat shock, we first examined how this leaky FLP activity would affect the outcomes of our color-based lineage analysis. From the non-heat shock experiments, our lineage-relatedness test determined that all neurons appear to be in distinct lineages (Figure 4, left panel). This indicates that the leaky FLP activity is unlikely to recombine the nBitbow cassette in any SOP until accumulation to a sufficiently high level to recombine in the post-mitotic neurons. This also suggests that our lineage-relatedness test is reliable to avoid the false-positive assignment of two cells being lineage related.

Previous studies used heat shock to sparsely induce lacZ expression at early developmental time points and referred cell clusters that are located nearby to be in the same SOP lineages (Brewster and Bodmer, 1995, 1996). We started nBitbow lineage tracing with 30-min heat shock at 3–4 h AEL. The lineage-relatedness test led to some of the same conclusions as previous studies: 4 distinct lineages give rise to a pair of ldaA-lesA, ldaB-lesB, v'ada-v'esB, or v'pda-v'es2 neurons, respectively (Figure S1).

The lacZ-based lineage tracing approach may incorrectly determine neurons in different lineages as being in the same lineage if they are easily labeled at the same heat shock time point. However, nBitbow labels the lineage-unrelated neurons in distinct colors even if their progenitors are labeled at the same heat shock time point. For instance, prior studies suggested that some dendritic arborization (da) neurons in the dorsal multidendritic (dmd) cluster were generated from one SOP and those in the ventral multidendritic (vmd) cluster were generated from another SOP (Figure 3A, shaded clusters) (Brewster and Bodmer, 1995, 1996). In contrast, our results showed that none of the da neurons from either the dmd or the vmd cluster are lineage related (Figure S2). The lacZ-based lineage tracing approach may also incorrectly determine neurons in the same lineage as being in different lineages if they migrate away from each other. However, nBitbow relies on color identify, not spatial proximity, and thus avoids this type of error. For instance, the 5 external sensory (es) neurons (desC, lesC, v'esA, vesA, and vesB) had been previously concluded to be the only neurons generated from 5 SOPs. Our lineage-relatedness test revealed that each of the 5 es neurons was lineage related to either a da or a tracheal dendrite (td) neuron (ddaD-desC, ltd-lesC, vdaA-v'esA, vdaC-vesB, and vdaD-vesA) (Figures 3C and 3D). This finding is in agreement with the results mentioned earlier that dda or vda neurons in the dmd or vmd clusters are not lineage related to each other but are lineage related to the des or ves neurons, respectively (Figures 3C and 3D; Figure S2).

Altogether, these results demonstrate the feasibility of using nBitbow for lineage studies in the PNS of the *Drosophila* larva and offer a lineage map of all larval PNS neurons.

Estimating the Birth Timing of PNS Neurons by nBitbow

One inherent benefit of using nBitbow for lineage mapping is its efficiency. nBitbow does not require sparse labeling. Indeed, all samples expressing any FP colors were used in the analysis. This allowed us to efficiently map the complete PNS lineages using only 17–25 larvae for each heat shock time point.

Comparing experiments done with 3–4 h AEL heat shock to those done with 5–6 h AEL heat shock, we found that even though many more neurons were labeled (Figure 2A), the overall percentage of neuron pairs that were labeled in matching colors decreased from 28.80% to 23.98%, respectively. Consistently, the lineage-relatedness tests indicated that the ddaD-desC, v'pda-v'es2, v'td2-vchA/vchB/vesC, v'esA-vdaA, and vdaDvesA neuron pairs no longer appeared to be lineage related when heat shock was applied at 5–6 h AEL (Figure 4, middle and right panels). Although the frequency of color matching of the v'ada-v'esB pair or the vdaC-vesB pair dropped significantly in the experiments done with 5–6 h AEL heat shock, that of the ldaA-lesA pair or the two es neurons in the v'es2 group remained unchanged (Figure S3). These observations suggest that at least one of the neurons in the v'ada-v'esB or the vdaC-vesB lineage finishes its final cell division earlier than those in the ldaA-lesA or the v'es2 lineage.

Altogether, the birth timing of individual lineages can be estimated by comparing the color matching frequency between earlier and later heat shock time points.

nBitbow Is Suitable for Studying Lineages in the *Drosophila* CNS

The use of nBitbow for lineage studies is not restricted to *Drosophila* PNS. Similar to the PNS, the ventral nerve cord (VNC) of *Drosophila* CNS contains repeated segments, each of which contains about 30 neuroblasts giving rise to approximately 350 neurons (Bossing et al., 1996; Doe, 1992; Schmidt et al., 1997). Although there is no marker for the identification of every CNS neuron, markers that label subsets of cells, such as Gal4 drivers that express transgenes in specific cell types, allow the application of nBitbow to trace lineages within subsets of CNS neurons. Recent studies have identified a subset of larval interneurons, called Basin neurons (Basin-1, Basin-2, Basin-3, and Basin-4), that receive synaptic inputs from nociceptors and mechanosensors (Ohyama et al., 2015). These neurons are speculated to be lineage related, based on their similar cell body locations, morphologies, and projection patterns (Ohyama et al., 2015). Using nBitbow and the Basin-specific-driver R72F11-Gal4, we found that the four Basins typically expressed the same colors induced by 1-h heat shock at 3–4 h AEL and were significantly related to each other (Figure 5). These results demonstrate the feasibility of using nBitbow for lineage studies in the *Drosophila* CNS.

DISCUSSION

In this study, we introduced a digital version of Brainbow, termed nBitbow, for lineage tracing with enhanced efficiency and improved accuracy. Importantly, we developed statistical methods to determine the likelihood of two cells being lineage related based on their nBitbow colors. Applying nBitbow to trace neuronal lineages in the PNS of *Drosophila* larva, we confirmed and rejected previously determined lineages and discovered unknown lineages. Moreover, nBitbow revealed the proximate birth timing of many PNS neuronal lineages. In addition, we demonstrated the feasibility of using nBitbow for lineage studies in the *Drosophila* CNS.

nBitbow Enhances the Efficiency in Lineage Mapping

Mapping lineages by nBitbow is more efficient compared with previous methods. For example, a previous study examined more than 5,000 embryos to identify the lineage relationships in the *Drosophila* larval PNS. This is because the labeling needs to be extremely sparse to reliably determine the frequency of neighboring cells expressing a lacZ marker that was induced by heat shock in a single progenitor cell (Brewster and Bodmer, 1995, 1996). Using nBitbow, we achieved a similar goal with 40 animals (23 for heat shock at 3–4 h AEL and 17 for heat shock at 5–6 h AEL). Moreover, power analysis suggested, based on the color activation frequency with 30-min heat shock at 3–4 h AEL, that only 124 samples (e.g., hemisegments for segmentally repeated lineages) would be needed to lineage a set of 300 cells with 80% power at a 50% cell color match rate (Figure S4A, left panel; STAR Methods). In addition, taking advantage of each larval PNS neuron being identifiable by its cell body position and the morphology of proximal dendrites (Bodmer and Jan, 1987; Grueber et al., 2002; Jan and Jan, 1982), we combined neuronal morphology staining with nBitbow's multi-color nuclear labeling to assign each identified neuron to a lineage. With these technical improvements, we were able to determine the lineage relationships in all larval PNS neurons with a small number of animals.

nBitbow Improves the Accuracy of Lineage Mapping

The ability of nBitbow to generate many more unique colors overcomes previous methods' fundamental limitation in labeling accuracy. With limited color choices, previous monochromic or multicolor labeling methods rely on physical proximity when determining whether two or more cells belong to the same lineage. This may generate false-negative or false-positive results; for instance, cells that migrate far from their sisters may be missed as in distinct lineages, or cells located close to the cluster boundary may be included in the neighboring lineages. Our method minimizes these errors by providing reliable statistics to determine whether two cells are lineage related based solely on nBitbow color, but not on their physical locations. For instance, we discovered that even though da neurons in the vmd or dmd clusters are physically close to each other, they are all generated by distinct progenitor cells (Figure S2). Moreover, many of these da neurons were lineage related to a nonmultipolar neuron located farther away (Figures 3C and 3D). These findings support the notion that nBitbow improves the accuracy of lineage mapping.

nBitbow Reveals Proximate Birth Times of Cells

Initiating nBitbow labeling at different developmental time points permitted convenient birth-timing analysis in the context of lineage relationship among PNS neurons (Figure 4). We found that several PNS neuronal pairs that were identified as lineage related in experiments with an earlier heat shock time point were identified as lineage unrelated in experiments with a later heat shock time point. This suggests that none of the neurons in those lineages were born at the earlier time point, so they inherited the same colors from their mother SOPs when heat shock was applied. However, one of the neurons in any of these lineages came to complete its last cell division at the later time point, so its color is distinct from its sister neuron. Nevertheless, this method could not be used to determine the birth timing of those cells that are the sole neurons in their lineages. Because nBitbow expresses bright nuclear-targeted FPs, it is possible that the birth timing of such neurons can be revealed by time-lapse imaging in live tissues. As a preliminary test, we performed time-lapse imaging of CNS cells labeled by nBitbow and observed that single mother cells divided into two daughter cells expressing the same color as their mother (Figure S4B).

Comparisons with Other Multicolor Labeling Systems

Different multicolor labeling systems have advantages and limitations for specific applications (Kretzschmar and Watt, 2012; Weissman and Pan, 2015). This is true for nBitbow.

First, the availability of drivers for activating colors through recombination limits the applicability of each system. For example, dBrainbow, UAS-Brainbows, LoLLIbow, and MCFO use Cre recombinase to activate color selection, while Flybows, TIE-DYE, and our nBitbow use FLP recombinase; Raeppli uses FLP to activate an integrase, which in turn activates color selection (Boulina et al., 2013; Hadjieconomou et al., 2011; Hampel et al., 2011; Kanca et al., 2014; Nern et al., 2015; Richier and Salecker, 2015; Worley et al., 2013). Because Raeppli requires extra time to activate colors, it might not be suitable for studying lineages, at least those occurring at early developmental stages.

Second, the subcellular localization and brightness of FPs are appropriate for specific purposes. For examples, Flybows, UAS-Brainbows, LoLLIbow, and Raeppli use membrane-bound FPs, allowing the visualization of fine morphological structures, such as dendritic spines or synapse, and can be used for identifying cells in certain circumstances. By contrast, nBitbow and Raeppli use nuclear FPs. Although they do not reveal cell morphology, they generally produce brighter FP signals, eliminating the need for immunostaining. Moreover, cells expressing nuclear FPs are easier to score than those expressing membrane FPs in tissues with a high density of cells.

Third, the diversity of colors is important for successful application of multicolor labeling systems in large numbers of cells. The more color variants available, the less likely that two random cells will express the same color. dBrainbow, Flybows, UAS-Brainbows, LoLLIbow, and Raeppli provide 3 or 4 colors per construct and MCFO and TIE-DYE provide 7 colors, whereas nBitbow provides more than 20 colors per construct, making it a great candidate for lineage tracing of large numbers of cells.

Finally, the direct signal detection and stability of FPs affect suitability for live imaging. For example, although MCFO provides more colors than other Brainbow variants, it requires immunostaining to enhance FP signals, precluding its use for live imaging. In contrast, both nBitbow and LoLLIbow can be used for live imaging. nBitbow is probably more suitable for visualizing larger scales of lineage patterns and distribution because of its greater color diversity. Membrane-bound LoLLIbow is probably better for imaging fine details of cell morphology in a small number of cells because of its 3 color variants.

Applying nBitbow to Lineage Mapping in Other Tissues

One concern about applying nBitbow in other tissues is that although the nuclear Bitbow is sufficient for cell identification in the *Drosophila* PNS, because neurons can be readily identified by anti-horseradish peroxidase (HRP) staining, it might be difficult to use in tissues in which cells cannot be easily identified. The solution is to focus on mapping lineages within a subset of cells by using specific Gal4 drivers. In this work, we demonstrated the feasibility of this approach with 14 larvae by showing that the four Basin neurons are lineage related (Figure 5). Additional power analysis suggested that only 34 samples would be needed for determining lineage relationship of 300 cells, based on the color activation frequency with 1-h heat shock at 3–4 h AEL and the same criteria mentioned previously (Figure S4A, right panel). Although our CNS demonstration does not show the identity of individual Basins, additional cell-type-specific markers (e.g., Basin subtype-specific Gal4) (Ohyama et al., 2015) or UAS-membrane markers can be recombined with nBitbow to identify individual neurons. If a Gal4 driver is weak, increasing Gal4 copies or immunostaining FPs would enhance nBitbow signals. If the expression of a Gal4 is dynamic, a cell-type-specific FLP with temporal expression in the early developmental stage can be used in combination with a constitutive Gal4. Alternatively, other binary expression systems (e.g., LexA/LexAop and QF/QUAS) can be combined with nBitbow to serve as cell-type-specific markers. Besides the nuclear-localized Bitbow, a membrane Bitbow may be useful for cell identification.

New Opportunities Resulted from Lineage Mapping by nBitbow

The success of our PNS lineage mapping created new opportunities for studying other aspects of neuronal development related to lineage development. For instance, it lays the foundation for studying whether particular molecular mechanisms are specified in neurons arising from different lineages. These neuronal properties include morphological selection, electrophysiology, neurotransmission, and circuit connections. Furthermore, we expect these tools to greatly increase the efficiency and accuracy for studying *Drosophila* lineage development beyond the nervous system. These tools also bring the potential of combining lineage mapping with other fields of research to answer complex questions in developmental biology.

STAR★METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Bing Ye (bingye@umich.edu). All unique/stable reagents generated in this study are available from the Lead Contact without restriction.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

To generate UAS-nBitbow flies, the nBitbow construct cDNAs encoding the following FPs were used: mAmetrine, mTFP, mNeonGreen, mKusabira-Orange2, and tdKatushka2 (Ai et al., 2006, 2008; Sakaue-Sawano et al., 2008; Shcherbo et al., 2009; Shaner et al., 2013). Human histone 2B protein (hH2B) was fused in frame to the N terminus of individual FPs (hH2B-FP). Individual incompatible FRT sequence pairs (FRT^{F3}, FRT^{F14}, FRT^{545}, FRT^{F13}, or FRT^{5T2}) was then placed in the opposing direction on both ends of the hH2B-FP sequence (Cai et al., 2013; McLeod et al., 1986; Schlake and Bode, 1994; Turan et al., 2010; Volkert and Broach, 1986). And then each of the five FRT-H2B-FP-invFRT cassettes were assembled into the pJFRC-MUH backbone vector (Addgene, Inc.) by standard cloning method, so that the expression of each cassette is under the control of an upstream activation sequence (UAS) and an p10 poly-adenylation (p10pA) sequence (Pfeiffer et al., 2012). The final nBitbow targeting plasmid was assembled by Gibson assembly (Gibson et al., 2009) into the same pJFRC-MUH backbone and was integrated into *Drosophila* melanogaster genome docking site VK00027 (BestGene, Inc.) using the FC31 integrase-mediated transgenesis systems (Bateman et al., 2006; Bischof et al., 2007; Groth et al., 2004; Markstein et al., 2008; Venken et al., 2006).

Other stocks: hs-FLP[122] (Rangarajan et al., 1999); *elav*-Gal4 (BDSC, #8760); GMR72F11-Gal4 (BDSC, #39786).

METHOD DETAILS

nBitbow transgenic flies, UAS-nBitbow, was crossed with flies that carry heat-shock-inducible FLP[122] and a Gal4 driver (elav-Gal4 for PNS and R72F11-Gal4 for Basin neurons) (Pfeiffer et al., 2008), and raised at 23C for 2 days. To induce FP expression in the larval PNS, embryos were collected within consecutive 1 hr windows on day 3, and kept at 23C until heat-shocked at 2–3, 3–4, 4–5, 5–6, or 6–7 hr AEL. To induce FP expression in

Basin neurons, embryos were heat-shocked at 3–4 hr AEL. To induce FP expression in thoracic lineage clones in the larval CNS, 1st instar larvae were heat-shocked on 1 day AEL. The heat shock was performed by heating the embryo-containing agar-plate on the surface of a 70C heat-block for 1.5 min, followed by a 30-min (or 1-hr for Basin neurons) heat chamber incubation at 37C. The plate was then transferred back to the 23C incubator for the next 4 days.

Late 3rd instar larvae were dissected to make fillet preparations in 1× hemolymph-like 3.1 saline buffer (Feng et al., 2004), immediately followed by fixation in 4% paraformaldehyde in 1× PBS for 20 min at room temperature (RT) with gentle shaking. Larval fillets were transferred to 1.5 mL Eppendorf tubes wrapped in tinfoil to prevent possible photobleaching by ambient lights. The samples were washed 3 times with 0.1% Triton X-100 in 1× PBS, 15 min each at RT, on a nutator, and then incubated in 5% normal donkey serum containing 0.1% Triton X-100 in 1× PBS for 30 min at RT. Alexa Fluor 647 AffiniPure Goat Anti-Horseradish Peroxidase antibody (1:500, Jackson Immuno Research) was added to the samples for overnight incubation at 4C with gentle shaking. Samples were washed 3 times as described above. Fillets were placed on poly-L-lysine-coated coverslips and mounted with ProLong Diamond Antifade mountant (Invitrogen™). Samples were left dry in the dark overnight at RT and imaged by confocal imaging on the following day or kept at 4C for up to a week until imaging.

nBitbow samples were imaged with a Zeiss LSM780 laser scanning confocal microscope. PNS data were collected from abdominal segments 2 to 7 (A2-A7). Fiji with customized plugin scripts was used to process raw images for FP spectral unmixing (Schindelin et al., 2012). Individual FP signals assuming the shape of a nucleus in the soma were identified on a single z-plane. If a particular FP signal were observed in a neuron, it is recorded as an "1" in the color code; if that FP's signal is not observed, it is then recorded as a "0." Color code of each neuron was recorded manually to create a data sheet for each experimental condition, which were subsequently analyzed computationally and statistically.

For time-lapse imaging, embryos were collected as described above. Embryos were then manually dechorionated and incubated at 25C until heat-shock at 37C for 30-min at 4–5 hr AEL, followed by incubation at 18C for recovery. Subsequently, the embryos were placed on a poly-L-Lysine coated coverslip and covered by a drop of Halocarbon oil 700 (Sigma-Aldrich) to prevent drying. Time-lapse imaging started at 9–10 hr AEL with 30-min intervals for 12 hours. A 40x oil immersion lens (NA = 1.3) was used for imaging.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of lineage-relatedness—The goal of our statistical methods was to systematically identify neurons that share a common lineage versus those that do not share a common lineage. A key aspect of this test is the ability to score the lineage-relatedness for a pair of two neurons based on the color codes recorded from the nBitbow experiments. To do this, we implemented a modified version of the Cohen's Kappa (k) to quantify the degree of color-code agreement between two neurons, which is defined as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e},$$

(1)

where p_0 and p_e are the degree of color code agreement between the two neurons calculated by observation and by chance, respectively.

To calculate the κ value for each neuron pair, we first calculate the probability of a single neuron n expressing a particular color c, Pn, c:

$$P_{n,c} = \frac{M_c}{N},$$

(2)

where, M_c is the total observation that a neuron expresses color c, and N is the total number of neurons. As shown in Figure 2D, the chance that a neuron expressed any fluorescent color was relatively low (6% maximum). We therefore hypothesize that most of the null-color neurons were the result of no Flp recombination at all, instead of due to the same FP being recombined even times. As a result, we estimated the κ in the cases that at least one of the neurons in a comparing pair needs to express at least one FP to avoid the false identification of null-color expressing neurons being lineage related. Based on this assumption, we can estimate the p_e value between neurons i and j (i.e., $p_e^{i,j}$), in which at least one of them expresses at least one FP. The $p_e^{i,j}$ value can be calculated by summing the products of the $P_{n,c}$ for both neurons for all the non null colors (denoted from 1 to C) then dividing by all cases where at least one neuron had color $(1 - (P_{i,0}) \times (P_{j,0}))$ to account for our modified κ value ignoring double blank matches:

$$p_e^{i,j} = \frac{\sum_1^C (P_{i,c}) \times (P_{j,c})}{1 - (P_{i,0}) \times (P_{j,0})}.$$

(3)

To calculate the p_0 value of the corresponding neuron pair $p_0^{i,j}$, we again, only considered the cases in which at least one of the neurons expressed at least one FP:

$$p_0^{i,j} = \frac{N_{i,j}^c}{M_{i,j}},$$

(4)

where, $N_{i,j}^c$ is the total observations that both neurons expressed a same non null colors color and $M_{i,j}$ is the total observations that either neuron expressed a non null color.

Finally, we acquired the observed κ values for all the neuron pairs using their p_e and p_0 values following Equation 1. An analysis of this equation reveals that if any two neurons always have perfectly matched colors (i.e., $p_0 = 1$), they will have a $\kappa = 1$. If two neurons' colors are matched exactly as would be expected by chance (i.e., $p_0 = p_e$), then they will have a $\kappa = 0$. In this way, a higher κ value indicates a higher probability that the two neurons are labeled in the same color not by chance (i.e., lineage-related).

Statistical analysis of lineage relatedness—We designed a statistical test to determine whether the experimentally observed κ value of a neuron pair is significantly higher than that by chance. To estimate the κ value distribution from labeling a particular neuron pair by random colors, we established a randomization process, starting with generating a raw color matrix where every row represents an individual PNS hemi-segment, and each column represents a neuron within that hemi-segment. The κ values were calculated for all possible combinations of neuron pairs within the data. Next, the raw color matrix was randomized by shuffling the colors each neuron expressing any non-00000 null-color. This ensures that the neurons receiving a color in the randomized data will have also had a color in the real data. We reason that this randomization process is more accurate than a total randomization process because each neuron pair remains the same total chance to match colors as in the real experiments, in which the ambiguous 00000 null-color are excluded from being assigning as a matched color (see above). After the color randomization, the κ values were calculated for all possible combinations of neuron pairs within the data. This randomization process was repeated 100,000 times for each neuron pair to generate a distribution of κ values that were compared to the κ value obtained from the experimental observation. The number of times that the random κ value was greater than or equal to the observed κ value was counted and then divided by 100,000 to generate a p value to estimate whether the experimentally observed κ value is significantly larger than those calculated by random color assignment. Finally, the p values was adjusted for multiple hypothesis testing by calculating a false discovery rate (FDR) using the Python statsmodels package (https://www.statsmodels.org/dev/generated/ statsmodels.stats.multitest.multipletests.html). The Benjamini/Hochberg (non-negative) method was used for FDR evaluation (Benjamini and Hochberg, 1995).

It is important to run a separate test for each neuron pair because certain neurons expressed a color more often than others. This allowed us to normalize the high expressing neurons in a way that they did not just have a better κ value by chance.

The complete lineages map of PNS neurons was generated by the Cytoscape Software package (version 3.7.1., https://cytoscape.org/index.html).

Power analysis of lineage-relatedness test—To quantify the number of samples needed to robustly lineage a set of neurons, we performed a power analysis on our test. This analysis modeled a single neuron pair within a theoretical larger lineage. This single neuron pair was assigned a number of samples (N), a color distribution (C) and an above background match percentage (P). This above background match percentage corresponds to the effect size of this analysis and represents the chance that two neurons match due to their lineage relationship.

To model N neuron pairs from color distribution C with a match percentage P, we first randomly assigned N neurons as either "matching" or "non-matching" based on P. A random number from 0 to 1 was generated. If that number is less than P, then the neuron pair is said to be matching. If a neuron pair is assigned to be matching, a single color is selected from color distribution C. As C contains blanks, it is possible that both neurons are assigned the null-color even if they are matching. If a neuron pair is not assigned to be

matching, two colors are chosen from C and assigned to the neurons separately. This process was repeated N times to generate the theoretical dataset.

This dataset containing a single neuron pair with N neurons was analyzed as described above with one important exception: the $p_e^{i,j}$ value was assigned based on the color distribution C not based on the observed colors within the small subset. This prevented wild changes to the κ value with small sample numbers and is consistent with our modeling of a pair within a larger set of neurons. A p value was calculated for this individual neuron pair using the randomization protocol described above.

To calculate the minimum number of samples (n) to obtain a power (Pr) at a given p value cutoff (a) this modeling process was repeated 1000 times. The number of times the p value was below a was divided by 1000 to obtain a power for the parameters specified in the analysis. If this power is above the desired Pr value, N is reduced. If the power is below the desired Pr, N is increased until the minimum number of samples required for the desired power is found (n).

For our analysis, we started with the 3–4 hr AEL color distribution as that is the most sparsely labeled and therefore the most conservative. We targeted $Pr = 80\%$ and calculated the minimum number of samples required to identify lineages with different effect sizes (match rates). The alpha level of our analysis was set at 0.000001, which corresponds to a 0.05 p value with a Bonferroni correction for all combinations of 300 cell pairs.

DATA AND CODE AVAILABILITY

The code for implementing the lineage-relatedness statistical model is available at https://github.com/MikeVeling/Process-Bitbow-Data

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

Ai HW, Henderson JN, Remington SJ, and Campbell RE (2006). Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural characterization and applications in fluorescence imaging. Biochem. J 400, 531–540. [PubMed: 16859491]

Ai HW, Hazelwood KL, Davidson MW, and Campbell RE (2008). Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors. Nat. Methods 5, 401–403. [PubMed: 18425137]

Bateman JR, Lee AM, and Wu CT (2006). Site-specific transformation of *Drosophila* via phiC31 integrase-mediated cassette exchange. Genetics 173, 769–777. [PubMed: 16547094]

Benjamini Y, and Hochberg Y (1995). Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B 57, 289–300.

Birkholz O, Rickert C, Nowak J, Coban IC, and Technau GM (2015). Bridging the gap between postembryonic cell lineages and identified embryonic neuroblasts in the ventral nerve cord of *Drosophila* melanogaster. Biol. Open 4, 420–434. [PubMed: 25819843]

Bischof J, Maeda RK, Hediger M, Karch F, and Basler K (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. USA 104, 3312–3317. [PubMed: 17360644]

Bodmer R, and Jan YN (1987). Morphological differentiation of the embryonic peripheral neurons in *Drosophila*. Rouxs Arch. Dev. Biol 196, 69–77.

Bodmer R, Carretto R, and Jan YN (1989). Neurogenesis of the peripheral nervous system in *Drosophila* embryos: DNA replication patterns and cell lineages. Neuron 3, 21–32. [PubMed: 2515889]

Bossing T, Udolph G, Doe CQ, and 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. [PubMed: 8873753]

Boulina M, Samarajeewa H, Baker JD, Kim MD, and Chiba A (2013). Live imaging of multicolor-labeled cells in *Drosophila*. Development 140, 1605–1613. [PubMed: 23482495]

Brewster R, and Bodmer R (1995). Origin and specification of type II sensory neurons in *Drosophila*. Development 121, 2923–2936. [PubMed: 7555719]

Brewster R, and Bodmer R (1996). Cell lineage analysis of the *Drosophila* peripheral nervous system. Dev. Genet 18, 50–63. [PubMed: 8742834]

Cachero S, and Jefferis GS (2011). Double brainbow. Nat. Methods 8, 217–218. [PubMed: 21358624]

Cai D, Cohen KB, Luo T, Lichtman JW, and Sanes JR (2013). Improved tools for the Brainbow toolbox. Nat. Methods 10, 540–547.

Cohen J (1960). A Coefficient of Agreement for Nominal Scales. Educ. Psychol. Meas 20, 37–46.

Doe CQ (1992). Molecular markers for identified neuroblasts and ganglion mother cells in the *Drosophila* central nervous system. Development 116, 855–863. [PubMed: 1295739]

Dwass M (1957). Modified Randomization Tests for Nonparametric Hypotheses. Ann. Math. Stat 28, 181–187.

Feng Y, Ueda A, and Wu CF (2004). A modified minimal hemolymph-like solution, HL3.1, for physiological recordings at the neuromuscular junctions of normal and mutant *Drosophila* larvae. J. Neurogenet 18, 377–402. [PubMed: 15763995]

Garcıá-Moreno F, Vasistha NA, Begbie J, and Molna r, Z. (2014). CLoNe is a new method to target single progenitors and study their progeny in mouse and chick. Development 141, 1589–1598. [PubMed: 24644261]

Ghysen A, and O'Kane C (1989). Neural enhancer-like elements as specific cell markers in *Drosophila*. Development 105, 35–52. [PubMed: 2509189]

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, and Smith HO (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. [PubMed: 19363495]

Golic KG, and Lindquist S (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. Cell 59, 499–509. [PubMed: 2509077]

Groth AC, Fish M, Nusse R, and Calos MP (2004). Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. Genetics 166, 1775–1782. [PubMed: 15126397]

Grueber WB, Jan LY, and Jan YN (2002). Tiling of the *Drosophila* epidermis by multidendritic sensory neurons. Development 129, 2867–2878. [PubMed: 12050135]

Hadjieconomou D, Rotkopf S, Alexandre C, Bell DM, Dickson BJ, and Salecker I (2011). Flybow: genetic multicolor cell labeling for neural circuit analysis in *Drosophila* melanogaster. Nat. Methods 8, 260–266. [PubMed: 21297619]

Hampel S, Chung P, McKellar CE, Hall D, Looger LL, and Simpson JH (2011). *Drosophila* Brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. Nat. Methods 8, 253–259. [PubMed: 21297621]

Hartenstein V, and Campos-Ortega JA (1984). Early neurogenesis in wild-type*Drosophila* melanogaster. Wilehm Roux Arch. Dev. Biol 193, 308–325. [PubMed: 28305340]

Hartenstein V, Rudloff E, and Campos-Ortega JA (1987). The pattern of proliferation of the neuroblasts in the wild-type embryo of *Drosophila* melanogaster. Rouxs Arch. Dev. Biol 196, 473–485.

Hobert O, and Kratsios P (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. Curr. Opin. Neurobiol 56, 97–105. [PubMed: 30665084]

Jan LY, and Jan YN (1982). Antibodies to Horseradish Peroxidase as Specific Neuronal Markers in *Drosophila* and in Grasshopper Embryos. Proc. Natl. Acad. Sci. USA 79, 2700–2704. [PubMed: 6806816]

Kanca O, Caussinus E, Denes AS, Percival-Smith A, and Affolter M (2014). Raeppli: a whole-tissue labeling tool for live imaging of *Drosophila* development. Development 141, 472–480. [PubMed: 24335257]

Kretzschmar K, and Watt FM (2012). Lineage tracing. Cell 148, 33–45. [PubMed: 22265400]

Lacin H, Chen HM, Long X, Singer RH, Lee T, and Truman JW (2019). Neurotransmitter identity is acquired in a lineage-restricted manner in the Drosophila CNS. eLife 8, e43701. [PubMed: 30912745]

Lee T (2017). Wiring the *Drosophila* Brain with Individually Tailored Neural Lineages. Curr. Biol 27, R77–R82. [PubMed: 28118595]

Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, and Lichtman JW (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature 450, 56–62. [PubMed: 17972876]

Loulier K, Barry R, Mahou P, Le Franc Y, Supatto W, Matho KS, Ieng S, Fouquet S, Dupin E, Benosman R, et al. (2014). Multiplex cell and lineage tracking with combinatorial labels. Neuron 81, 505–520. [PubMed: 24507188]

Markstein M, Pitsouli C, Villalta C, Celniker SE, and Perrimon N (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat. Genet 40, 476–483. [PubMed: 18311141]

McLeod M, Craft S, and Broach JR (1986). Identification of the crossover site during FLP-mediated recombination in the Saccharomyces cerevisiae plasmid 2 microns circle. Mol. Cell. Biol 6, 3357–3367. [PubMed: 3540590]

Nern A, Pfeiffer BD, and Rubin GM (2015). Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. Proc. Natl. Acad. Sci. USA 112, E2967–E2976. [PubMed: 25964354]

Nichols TE, and Holmes AP (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp 15, 1–25. [PubMed: 11747097]

Nichols T, and Holmes A (2007). Non-parametric procedures. In Statistical Parametric Mapping: The Analysis of Functional Brain Images, Friston K, Ashburner J, Kiebel S, Nichols T, and Penny W, eds. (Academic Press), pp. 253–272.

Ohyama T, Schneider-Mizell CM, Fetter RD, Aleman JV, Franconville R, Rivera-Alba M, Mensh BD, Branson KM, Simpson JH, Truman JW, et al. (2015). A multilevel multimodal circuit enhances action selection in *Drosophila*. Nature 520, 633–639. [PubMed: 25896325]

Pan YA, Freundlich T, Weissman TA, Schoppik D, Wang XC, Zimmerman S, Ciruna B, Sanes JR, Lichtman JW, and Schier AF (2013). Zebrabow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish. Development 140, 2835–2846. [PubMed: 23757414]

Papaioannou VE (2016). Concepts of Cell Lineage in Mammalian Embryos. Curr. Top. Dev. Biol 117, 185–197. [PubMed: 26969978]

Pfeiffer BD, Jenett A, Hammonds AS, Ngo TTB, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Laverty TR, et al. (2008). Tools for neuroanatomy and neurogenetics in *Drosophila*. Proc. Natl. Acad. Sci. USA 105, 9715–9720. [PubMed: 18621688]

Pfeiffer BD, Truman JW, and Rubin GM (2012). Using translational enhancers to increase transgene expression in *Drosophila*. Proc. Natl. Acad. Sci. USA 109, 6626–6631. [PubMed: 22493255]

Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, and Schier AF (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat. Biotechnol 36, 442–450. [PubMed: 29608178]

Rangarajan R, Gong Q, and Gaul U (1999). Migration and function of glia in the developing *Drosophila* eye. Development 126, 3285–3292. [PubMed: 10393108]

Richier B, and Salecker I (2015). Versatile genetic paintbrushes: Brainbow technologies. Wiley Interdiscip. Rev. Dev. Biol 4, 161–180. [PubMed: 25491327]

Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, Kashiwagi S, Fukami K, Miyata T, Miyoshi H, et al. (2008). Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell 132, 487–498. [PubMed: 18267078]

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. [PubMed: 22743772]

Schlake T, and Bode J (1994). Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry 33, 12746–12751. [PubMed: 7947678]

Schmidt H, Rickert C, Bossing T, Vef O, Urban J, and 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. [PubMed: 9299113]

Schmidt ST, Zimmerman SM, Wang J, Kim SK, and Quake SR (2017). Quantitative Analysis of Synthetic Cell Lineage Tracing Using Nuclease Barcoding. ACS Synth. Biol 6, 936–942. [PubMed: 28264564]

Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M, et al. (2013). A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat. Methods 10, 407–409. [PubMed: 23524392]

Shcherbo D, Murphy CS, Ermakova GV, Solovieva EA, Chepurnykh TV, Shcheglov AS, Verkhusha VV, Pletnev VZ, Hazelwood KL, Roche PM, et al. (2009). Far-red fluorescent tags for protein imaging in living tissues. Biochem. J 418, 567–574. [PubMed: 19143658]

Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, and Junker JP (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nat. Biotechnol 36, 469–473. [PubMed: 29644996]

Stent GS (1985). The role of cell lineage in development. Philos. Trans. R. Soc. Lond. B Biol. Sci 312, 3–19. [PubMed: 2869528]

Truman JW, and Bate M (1988). Spatial and temporal patterns of neurogenesis in the central nervous system of *Drosophila* melanogaster. Dev. Biol 125, 145–157. [PubMed: 3119399]

Turan S, Kuehle J, Schambach A, Baum C, and Bode J (2010). Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. J. Mol. Biol 402, 52–69. [PubMed: 20650281]

Venken KJ, He Y, Hoskins RA, and Bellen HJ (2006). P[acman]: A BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science 314, 1747–1751. [PubMed: 17138868]

Volkert FC, and Broach JR (1986). Site-specific recombination promotes plasmid amplification in yeast. Cell 46, 541–550. [PubMed: 3524855]

Weissman TA, and Pan YA (2015). Brainbow: new resources and emerging biological applications for multicolor genetic labeling and analysis. Genetics 199, 293–306. [PubMed: 25657347]

Woodworth MB, Girskis KM, and Walsh CA (2017). Building a lineage from single cells: genetic techniques for cell lineage tracking. Nat. Rev. Genet 18, 230–244. [PubMed: 28111472]

Worley MI, Setiawan L, and Hariharan IK (2013). TIE-DYE: a combinatorial marking system to visualize and genetically manipulate clones during development in *Drosophila* melanogaster. Development 140, 3275–3284. [PubMed: 23785055]

Younossi-Hartenstein A, and Hartenstein V (1997). Pattern, time of birth, and morphogenesis of sensillum progenitors in *Drosophila*. Microsc. Res. Tech 39, 479–491. [PubMed: 9438249]

Highlights

- A single nBitbow cassette generates up to 31 color variants

- nBitbow enhances the efficiency and improves the accuracy of cell lineage tracing

- nBitbow is suitable for lineage tracing in both the PNS and the CNS of *Drosophila*

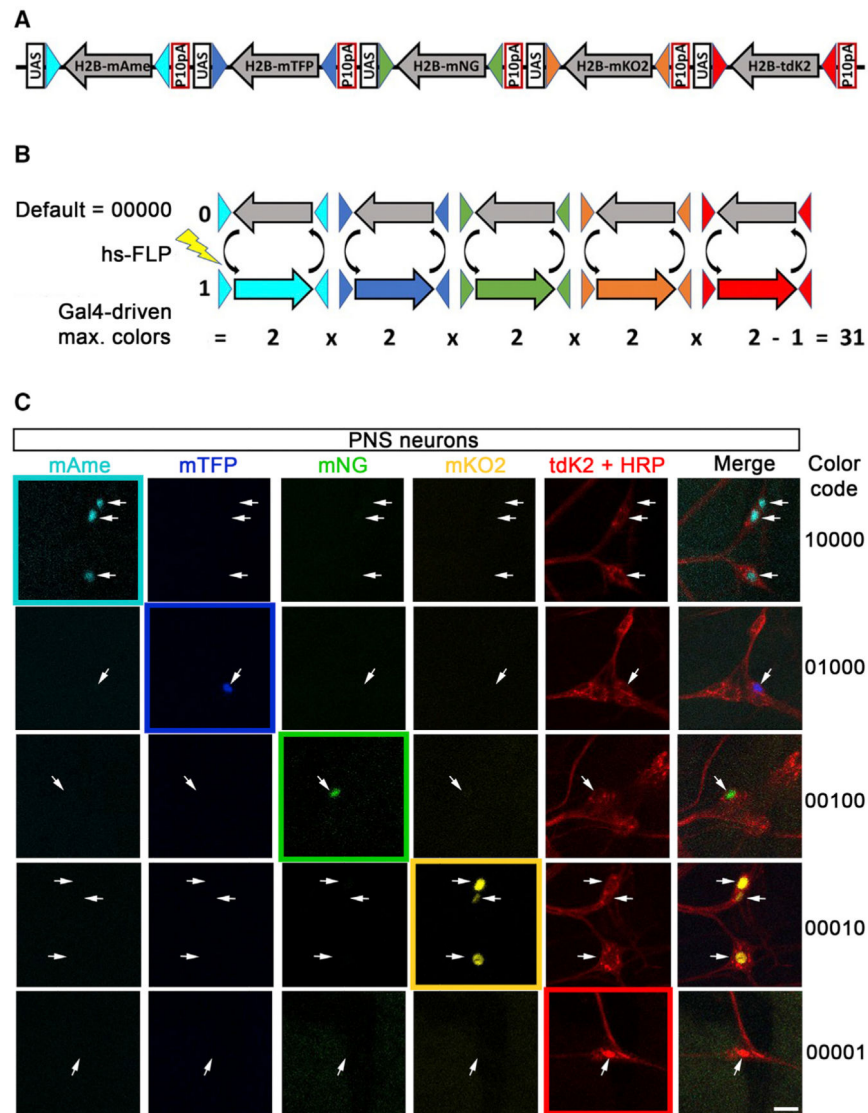- The lineage relationship of all PNS neurons in *Drosophila* larvae is determined

Figure 1. Design and Color-Coding Principles of nBitbow

(A) Schematic of the nBitbow spinning design The H2B fusion nuclear FP (nucFP) open reading frames are positioned in the reverse direction as a default non-detectable color in the absence of FLP.

(B) Schematics of color-coding in nBitbow. Upon heat shock-induced FLP expression, which causes recombination of FRT sites, each H2B-FP module can stochastically and independently spin into a forward (1, visible) or reversed (0, nondetectable) expression state. A maximum of 31 visible colors can be detected in the presence of a Gal4 driver.

(C) Spectral imaging and linear unmixing are performed to identify the five-bit nBitbow color code for each cell, in which 1s represent the FPs that are turned on and 0s represent those that are not turned on. Images shown are PNS neurons. HRP, anti-HRP immunostaining, which specifically labels PNS neurons. Scale bar: 10 mm.
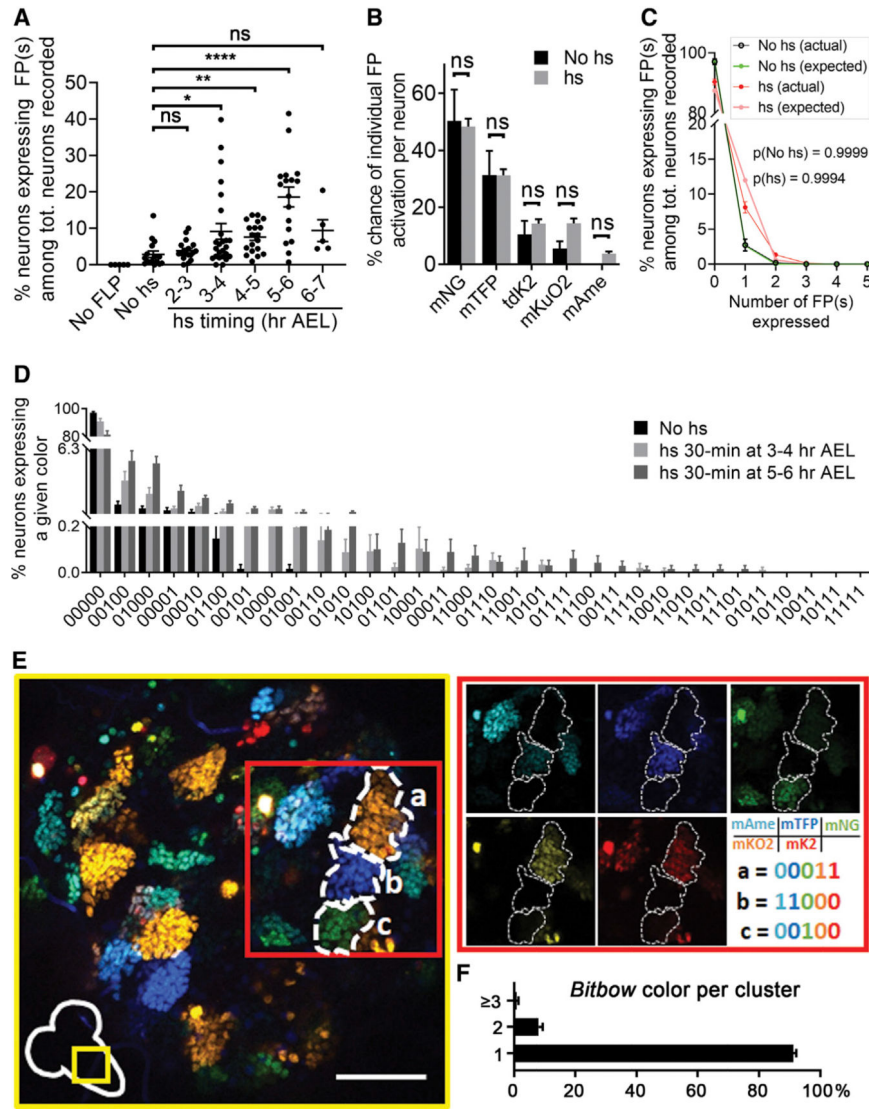
Figure 2. Labeling and Lineage Tracing Efficacy of nBitbow in the *Drosophila* Nervous System
(A) Quantification of the labeling coverage of nBitbow by heat shock FLP induction in the PNS at various time points during early embryonic development. Sample numbers (larvae): no FLP, 5; no hs, 16; 2–3 h AEL, 18; 3–4 h AEL, 25; 4–5 h AEL, 19; 5–6 h AEL, 17; 6–7 h AEL, 5. The p values were calculated by Dunn's multiple comparison test. ns, $p > 0.05$; *$0.01 < p \leq 0.05$; **$0.001 < p \leq 0.01$; ***$0.001 < p \leq 0.001$; and ****$p < 0.0001$.
(B) Comparison of the turn-on efficiency of each FRT-nucFP-FRT module in the PNS. Heat shock (hs) data are the combined data from hs at 2–3, 3–4, 4–5, 5–6, and 6–7 h AEL. Sample numbers (larvae): no hs, 16; hs, 81. The p values were calculated by multiple comparison using the Holm-Sidak method.
(C) The actual FP frequency distribution is not different from the theoretical distribution (expected) as the number of activated FP increases in the PNS. This indicates that the recombination of each FRT-nucFP-FRT module is independent of each other. Sample numbers (larvae): no hs, 15; hs, 81. The p values were calculated by the chi-square goodness-of-fit test for Poisson distribution.

(D) 27 color codes were observed with different frequencies in the PNS from a *Drosophila* line that carries only one nBitbow cassette. Sample numbers (larvae): no hs, 15; 3–4 h AEL, 25; 5–6 h AEL, 17.

(E) Representative images of nBitbow labeling in thoracic segments in the CNS of the third-instar larvae with embryonic heat shock FLP induction. Most cell clusters (i.e., lineage clones) were labeled with the same color. 193 clusters were analyzed. Scale bar: 50 μm.

(F) Proportion of thoracic lineage clones labeled by one, two, or more Bitbow color tags, quantified across n = 193 clusters. The p values were calculated using unpaired two-tailed t test, and α of 0.05 is used as the cutoff for significance. Error bars: SEM.
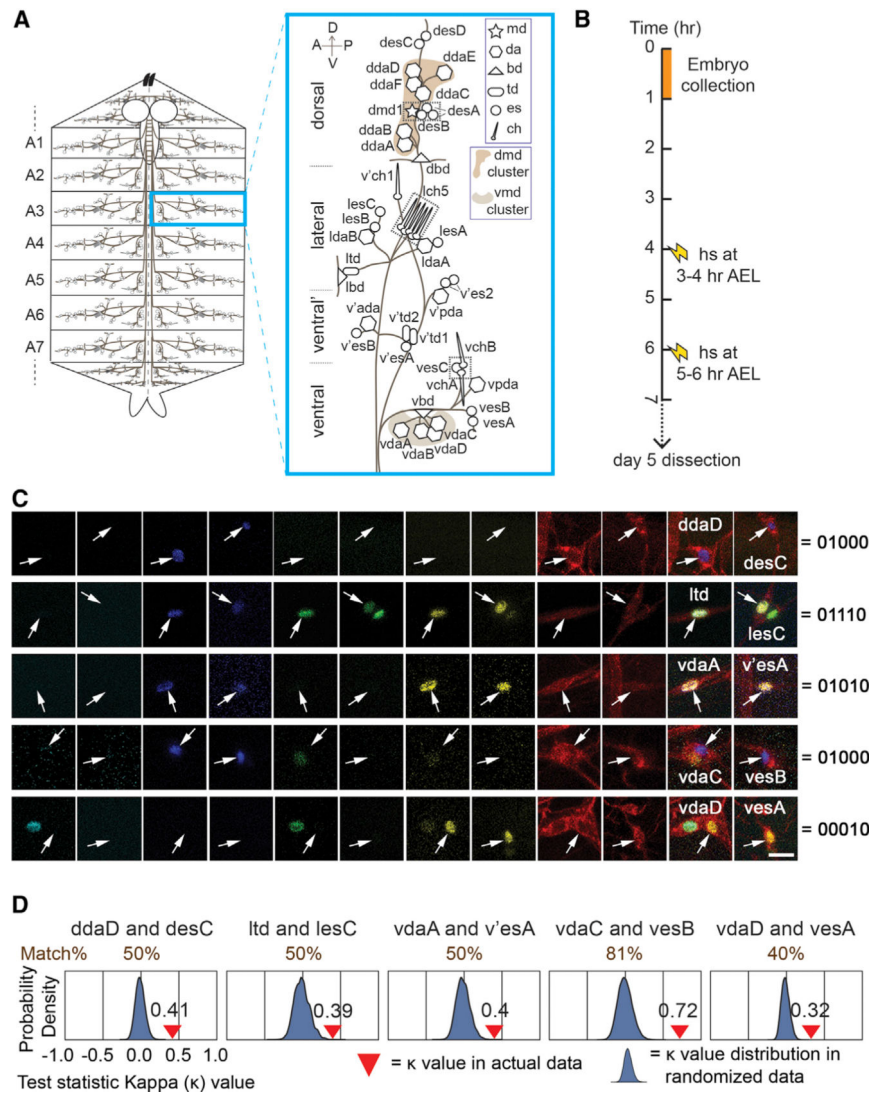
Figure 3. nBitbow Efficiently Identifies Neuronal Lineages of *Drosophila* PNS Neurons
(A) Schematic drawing of the distribution of PNS neurons in one hemisegment of the larval body wall. In the PNS, multipolar neurons include the md, da, bipolar dendrite (bd), and td types, whereas bipolar neurons include the es and chordotonal (ch) types.
(B) Schematic drawing of the experimental steps.
(C) Representative fluorescent images that show matching color codes of pairs of neurons previously thought to be independent of each other. These neuronal pairs are the five es neurons (desC, lesC, v'esA, vesA, and vesB) and their lineage-related da or td neurons. These es neurons have been previously reported as the only neuron from individual SOPs (Brewster and Bodmer, 1995). Larvae were heat-shocked at 3–4 h AEL. Scale bar: 10 μm.
(D) Statistical analysis confirms that these neuron pairs are lineage related. The κ values calculated based on experimental observations (red arrows) are significantly higher (i.e., closer to 1) than those calculated based on a random color distribution (blue).
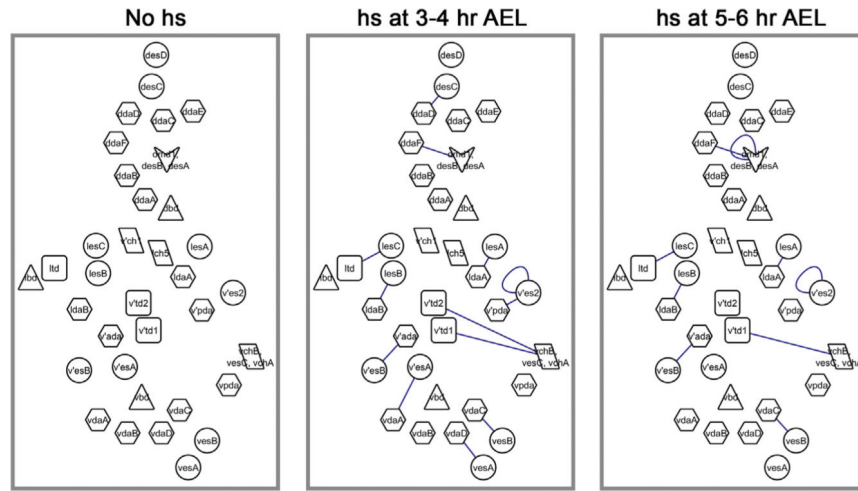See also Figures S1 and S2.

Figure 4. PNS Lineage Maps Revealed by nBitbow at Different Developmental Times
Neurons in a hemisegment are shown. These maps are based on the lineage relationships of all neurons determined by their calculated κ values from experiments with no heat shock (left), heat shock at 3–4 h AEL (middle), and heat shock at 5–6 h AEL (right). Each line connects a pair of lineage-related neurons with 2% false discovery rate (FDR). Three groups of neurons that cannot be identified individually (dotted boxes in Figure 3A) are presented as single symbols in these maps: the arrowhead includes four neurons—dmd1, desB, and 2 desA—that are located close together; the parallelogram labeled as lch5 includes 5 ch neurons; and the parallelogram labeled as vchA, vchB, and vesC includes the three neurons indicated.
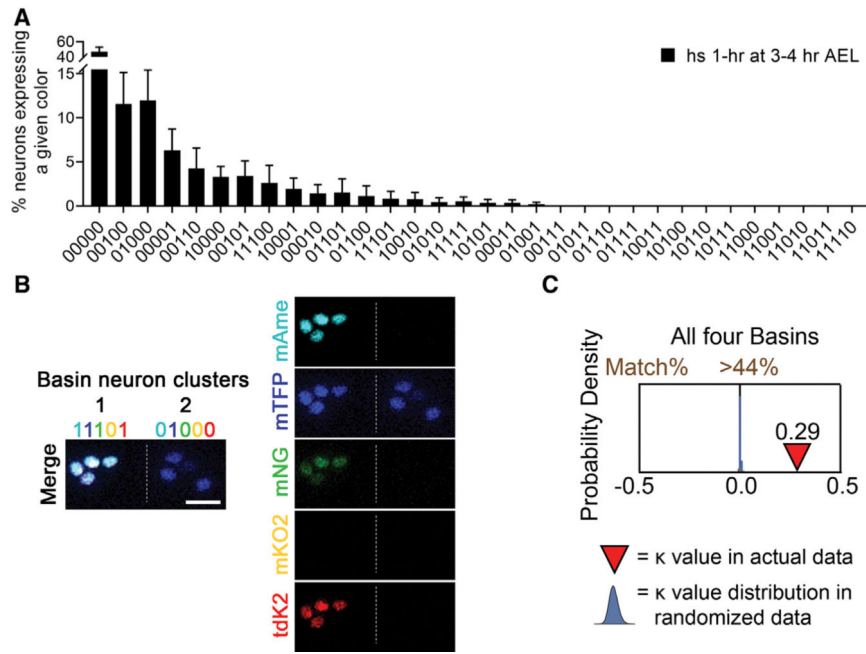See also Figure S3.

Figure 5. Lineage Tracing of Basin Neurons in the CNS of *Drosophila*

(A) Quantification of color activation frequency inBasin neurons with 1-h heat shock at 3–4 h AEL. Sample number (larvae): 13. Error bars: SEM.

(B) Representative fluorescent images that showBasin neurons expressing the same color in a cluster that is different from the adjacent cluster. Scale bar: 10 μm.

(C) κ value and random distribution of Basin neurons. Statistical analysis confirms that all four basin neurons are lineage related.

See also Figure S4

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Alexa Fluor 647 AffiniPure Goat Anti-Horseradish Peroxidase antibody	Jackson Immuno Research	Cat# 123–605-021; RRID: AB_2338967
Experimental Models: Organisms/Strains		
D. melanogaster: yw;;UAS-nBitbow	This paper	N/A
D. melanogaster: hs-flp[122];;	Rangarajan et al., 1999	N/A
D. melanogaster: w*;;elav-Gal4	Bloomington *Drosophila* Stock Center	RRID:BDSC_8760
D. melanogaster: w[1118];;GMR72F11-Gal4	Bloomington *Drosophila* Stock Center	RRID:BDSC_39786
Recombinant DNA		
Plasmid: pJRFC-MUH-nBitbow	This paper	N/A
Software and Algorithms		
Fiji	NIH	RRID:SCR_002285
Cytoscape (version 3.7.1)	NIGMS	RRID:SCR_003032
Python Statsmodels	NIDA	RRID:SCR_016074
ZEN (2012 SP1 black edition, release 8.1.0)	Carl Zeiss Microscopy	RRID:SCR_013672
Leica Microsystems LAS AF (version 2.6.3)	Leica Microscopy	RRID:SCR_008960
Prism 8	GraphPad	RRID:SCR_002798
Process-Bitbow-Data	This Paper	https://github.com/MikeVeling/Process-Bitbow-Data