Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing

Owen J Marshall[1–4], Tony D Southall[1–4], Seth W Cheetham[1,2] & Andrea H Brand[1,2]

[1]The Gurdon Institute, University of Cambridge, Cambridge, UK. [2]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. [3]Present addresses: Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia (O.J.M.), Department of Life Sciences, Imperial College London, London, UK (T.D.S.). [4]These authors contributed equally to this work. Correspondence should be addressed to A.H.B. (a.brand@gurdon.cam.ac.uk).

Published online 4 August 2016; doi:10.1038/nprot.2016.084
This protocol is an extension to *Nat. Protoc.* **2**, 1467–1478 (2007); doi:10.1038/nprot.2007.148; published online 7 June 2007

The ability to profile transcription and chromatin binding in a cell-type-specific manner is a powerful aid to understanding cell-fate specification and cellular function in multicellular organisms. We recently developed targeted DamID (TaDa) to enable genome-wide, cell-type-specific profiling of DNA- and chromatin-binding proteins *in vivo* without cell isolation. As a protocol extension, this article describes substantial modifications to an existing protocol, and it offers additional applications. TaDa builds upon DamID, a technique for detecting genome-wide DNA-binding profiles of proteins, by coupling it with the GAL4 system in *Drosophila* to enable both temporal and spatial resolution. TaDa ensures that Dam-fusion proteins are expressed at very low levels, thus avoiding toxicity and potential artifacts from overexpression. The modifications to the core DamID technique presented here also increase the speed of sample processing and throughput, and adapt the method to next-generation sequencing technology. TaDa is robust, reproducible and highly sensitive. Compared with other methods for cell-type-specific profiling, the technique requires no cell-sorting, cross-linking or antisera, and binding profiles can be generated from as few as 10,000 total induced cells. By profiling the genome-wide binding of RNA polymerase II (Pol II), TaDa can also identify transcribed genes in a cell-type-specific manner. Here we describe a detailed protocol for carrying out TaDa experiments and preparing the material for next-generation sequencing. Although we developed TaDa in *Drosophila*, it should be easily adapted to other organisms with an inducible expression system. Once transgenic animals are obtained, the entire experimental procedure—from collecting tissue samples to generating sequencing libraries—can be accomplished within 5 d.

INTRODUCTION

What makes cells different from each other and how their different properties are specified are key questions for developmental biologists, physiologists and neurobiologists. Profiling gene expression patterns and transcriptional networks in a cell-type-specific manner is a powerful method for investigating the mechanisms that specify cell fate and cell properties.

Development of TaDa

We recently developed 'TaDa' to profile genome-wide, cell-type-specific protein binding *in vivo* without cell isolation[1]. TaDa does not require the purification of cells or nuclei, and the technique is simple, requiring no fixation or immunoprecipitation. TaDa is based on DamID (DNA adenine methyltransferase identification)[2,3], which is an *in vivo* chromatin profiling technique. For DamID, *Escherichia coli* DNA adenine methyltransferase (Dam) protein is fused to a DNA- or chromatin-binding protein of interest. When the fusion protein is expressed *in vivo*, its binding sites are tagged by adenine methylation of the sequence GATC, enabling DNA fragments containing the sites to be isolated after enzymatic digestion. However, expression of Dam at high levels is toxic, and as a result DamID requires the Dam-fusion protein to be expressed at extremely low levels[2]. A drawback of this approach is that the Dam-fusion protein is expressed constitutively from basal promoters in all cell types. Driving expression of Dam fusions using targeted expression systems, such as the GAL4 system[4], results in excessively high levels of methylase and toxicity[1]. As a result, it was previously impossible to perform cell-type-specific DamID *in vivo*.

Therefore, we originated TaDa to reduce the level of translation of the Dam-fusion protein and to enable expression driven by the GAL4 system or other targeted expression systems. TaDa achieves this by using an unusual feature of eukaryotic ribosome translation whereby translation of a secondary open reading frame (ORF) present on a transcript can be initiated, but at a greatly reduced efficiency. The frequency of translational initiation at the secondary ORF is inversely proportional to the length of the primary ORF, allowing very fine control over protein translation levels[1,5,6]. We have shown that TaDa fusion proteins can be driven in the *Drosophila* CNS throughout development and adult life without toxicity. Importantly, TaDa enables transcription factors to be expressed in a cell-type-specific manner at extremely low levels, to allow profiling without altering cell fate (**Fig. 1**).

Overview of the procedure

The TaDa procedure involves two main phases (**Fig. 2**). Briefly, in the first phase, tissue is collected and genomic DNA is extracted (Steps 1–11) and digested with the restriction enzyme DpnI, which cuts only at adenine-methylated GATC sites. The digested DNA is column-purified to exclude uncut, genomic DNA from uninduced cells (Steps 12–14), and adaptors for PCR amplification are ligated to the DpnI cut fragments (Steps 15–18). To prevent unmethylated regions of DNA from being aberrantly amplified, the material is digested with DpnII, which cuts only non-methylated GATC sites (Steps 19–21; **Table 1**). The resulting DNA is then used as a template for a PCR to amplify the methylated fragments (Steps 22–25). The product of the PCR will thus

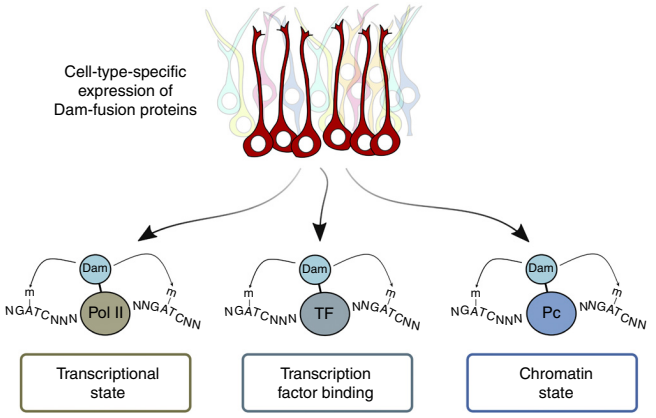

Figure 1 | TaDa can profile the interaction of any protein with DNA or chromatin in a cell-type-specific manner. For example, Pol II can be profiled to assess genome-wide transcription. TaDa can also profile transcription factor (TF) binding and assess chromatin states, by profiling chromatin-modifying enzymes such as Polycomb (Pc).

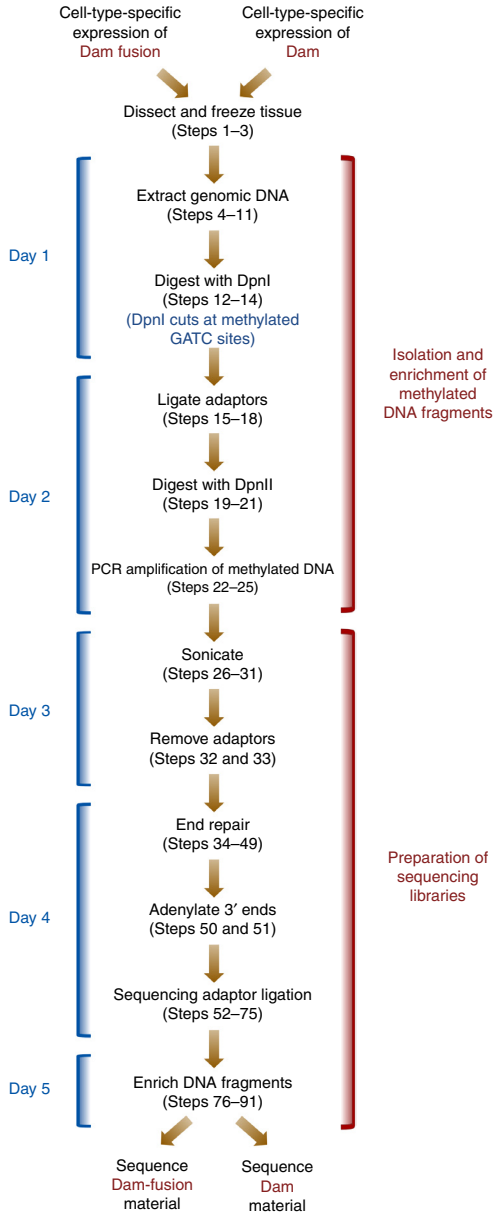

Figure 2 | Flowchart for processing of TaDa material, with protocol steps listed. The entire procedure can be completed in <5 d.

be highly enriched for bound GATC fragments. Although based on the original DamID technique[3], the procedure described here makes a number of modifications to the method, including avoiding DNA precipitation steps, removing unmethylated genomic material from uninduced cells (via column purification) and increasing the speed of the procedure.

In the second phase, the DNA is purified and then sonicated in preparation for next-generation sequencing (NGS; Steps 26–30), as some DNA fragments generated by the DamID procedure are too large for sequencing. To increase the mapping specificity of reads and to overcome problems of low diversity, DamID adaptors are removed with AlwI (Steps 31–33), which also removes the initial low-diversity GATC sequence at the start of amplified fragments; given the large size of many GATC fragments and the initial low diversity, NGS adaptors (**Supplementary Table 1**) cannot be directly used for the initial enrichment step in phase 1. NGS libraries are then constructed using a custom procedure (Steps 34–88; a commercial kit can be substituted here if required) in which the number of library PCR cycles is reduced (Steps 89–93). After Illumina sequencing, the binding profile of the fusion protein is determined from the sequencing data using an open-source and freely available software pipeline, and either peaks or transcribed genes (in the case of RNA Pol II) are identified (**Box 1**).

Applications and future uses of the method

We have successfully used TaDa to profile transcriptional activity, transcription factor binding and chromatin factor binding in multiple cell and tissue types in *Drosophila melanogaster*, including embryonic and larval neural stem cells, and larval and adult neurons (refs. 1,7; data not shown). To assess transcription, we fused Dam to RNA Pol II, thus enabling the identification of transcribed genes in a cell-type-specific manner[1]. Dam–Pol II can be used to compare the transcriptional states of closely related cell types. It can also be used to assess cell-type-specific changes in transcription in a mutant background, e.g., by simultaneously profiling with Dam–Pol II while knocking down a gene using RNAi (P. Speder and A.H.B., data not shown). Transcription factors are natural candidates for the technique, and their binding sites can also be profiled in a cell-type-specific manner (P. Fox and A.H.B., data not shown). Chromatin states can be broadly classified into five different types, which can be identified by profiling five different proteins with DamID[8]. We have mapped the binding of these five factors by TaDa, thereby enabling cell-type-specific profiling of chromatin states in the developing organism (O.M., J. Ander and A.H.B., data not shown). The technique is highly robust, and the sole limiting factor to cell-type-specific profiling is the availability of specific GAL4 driver lines.

DamID has been used successfully in yeast[9], plants[10], *Caenorhabditis elegans*[11] and human tissue culture cells[12,13]. Therefore, TaDa should be amenable for use in other eukaryotic organisms when coupled with appropriate cell-type-specific control of induction. Our preliminary data suggest that TaDa works successfully in mammalian cells (S. Cheetham, W. Gruhn, A. Surani and A.H.B., data not shown).

Box 1 | Processing of sequencing data ● TIMING 60 min per sample

Because the final form of DamID data is a $\log_2$ ratio of [Dam-fusion/Dam], care is required to ensure correct normalization between the fusion protein and the Dam-only control. Both Dam alone and Dam-fusion proteins will methylate highly accessible regions of the genome when localized to the nucleus, and this background methylation is used as the basis of normalization.

We have developed the damidseq_pipeline[7], an automated pipeline for processing samples that is freely available online (http://owenjm.github.io/damidseq_pipeline/), along with detailed usage instructions. The pipeline handles automatic alignment, extension, normalization and background reduction of the generated sequencing reads; files for human, mouse and fly genomes are provided on the website, and utilities are provided to generate the required files for any other organism. The damidseq_pipeline may be run on files in FASTQ or BAM format, as provided by a sequencing facility, and it generates a final $\log_2$ ratio file in bedgraph format.

Following ratio file generation, peaks with an FDR < 0.01 can be identified using find_peaks (http://github.com/owenjm/find_peaks). In the specific case of RNA Pol II data sets, the transcriptome can be identified using polii.gene.call (http://github.com/owenjm/polii.gene.call), which identifies all genes with an enriched occupancy (implying transcription) and an FDR < 0.01.

An example of the final output, illustrating the DNA-binding profile of RNA Pol II in *Drosophila* larval neural stem cells[7], is shown in **Figure 4**.

Comparisons with other methods

A number of techniques currently exist for performing cell-type-specific profiling; these take one of two approaches. The first approach focuses on specifically labeling and isolating the subset of cells of interest before undertaking conventional ChIP profiling. Examples include BITS-ChIP, in which cells[14] or fixed nuclei[15,16] are isolated by fluorescence-activated cell sorting, and the INTACT method, in which nuclei are isolated by biotin labeling and pulldown[17–19]. These methods have the advantage that they allow a large range of factors to be profiled via a single transgenic animal. However, they have the disadvantage of requiring very large amounts of material (typically in excess of 10^6 cells for ChIP), they can be technically challenging and they can potentially incorporate unlabeled cells or nuclei.

The second approach is to label specific factors in a cell-type-specific manner, precluding the need for cell-type purification. For example, in the CAST-ChIP method[20], GFP-tagged fusion proteins are expressed under the control of the GAL4 system and then, using an anti-GFP antibody, profiled by ChIP. This method requires no cell-type-specific purification, but it involves expressing the protein of interest at high levels, which may in turn influence cell fate or result in aberrant fusion-protein binding. Cell-type-specific transcriptional profiling can also be performed in this manner[21], via the targeted expression of a tagged RNA-binding protein[22], a ribosomal protein[23] or an RNA-modifying enzyme[24]. These methods often require large amounts of starting material and immunoprecipitation.

By contrast, TaDa requires neither large numbers of dissociated or fixed cells nor highly induced protein expression levels. TaDa profiles protein binding *in vivo* without potential artifacts that may arise from fixation, and it does so with very low levels of protein translation (so low that TaDa fusion proteins are undetectable via western blotting[1]), thus minimizing the impact of protein overexpression. Furthermore, the number of induced cells required for TaDa profiling is extremely low: we have successfully transcriptionally profiled ~100 neurons in whole *Drosophila* heads (>200,000 cells per head, and thus a 1:2,000 ratio of methylated DNA to total DNA), using only 100 heads and thus ~10,000 cells in total (T.D.S. and A.H.B., data not shown).

Limitations of TaDa

The resolution of TaDa depends on the frequency of GATC sites in the genome (present in the *Drosophila* genome at a median spacing of ~190 bp). Therefore, it cannot always provide the same resolution as ChIP-sequencing experiments. TaDa profiling of Pol II occupancy does not distinguish between all alternatively spliced transcripts, and it cannot provide information on the direction of transcription. TaDa profiling of Pol II occupancy will provide qualitative differences in transcription between cell types or conditions; however, it does not provide quantitative levels of mRNA produced. Although TaDa cannot directly profile histone modifications, it can profile the proteins that generate, remove or recognize these modifications.

Experimental design

Dam-fusion protein expression vector. To obtain inducible expression coupled with low-level translation, DNA-binding proteins of interest should be cloned into the pUAST-attB-LT3-NDam vector[1] (**Supplementary Fig. 1**; the full vector sequence is available from GenBank (accession no. KU728166); vector available upon request to A.H.B.). The vector uses mCherry as the primary ORF in the bicistronic transcript, allowing easy confirmation of GAL4 induction. Coding sequences cloned in frame into the multiple cloning site (MCS) of the vector will create a fusion protein with Dam at the N terminus of the protein, followed by a myc tag and a short linker separating Dam and the protein of interest. It is important to note that, after cloning or re-transformation, the Dam coding sequence should always be sequenced. We have occasionally observed spontaneous mutations arising in the Dam sequence after passage through bacteria, possibly owing to the role of adenine methylation in mismatch repair in *E. coli*[25]. Bacteria carrying mutated Dam appear to have a growth advantage; therefore, we tend to select the smaller colonies after transformation of the plasmid.

Cell-type-specific expression. To profile protein–DNA interactions in cells of interest, the choice of the GAL4 driver line is important. TaDa is highly sensitive, and therefore it is important that the GAL4 line be specific to the cells of interest. Drivers generated by homologous recombination[26] or by conversion of MiMIC lines[27] are usually optimal. The split-GAL4 system[28] is also an option for targeting specific groups of cells that are not delimited by the expression of a single gene.

Temporal control of expression. To achieve temporal control of TaDa, a temperature-sensitive repressor of GAL4,

GAL80^{ts}, can be used[29,30]. GAL80^{ts} inactivates GAL4 at 18 °C but not at 29 °C. With GAL80^{ts} in the genetic background, TaDa profiling can be initiated at any point during development or adult life by rearing the organism at 18 °C and shifting to 29 °C as required. Constitutively expressed GAL80^{ts} (tub-GAL80^{ts}) can be introduced into the background of either the GAL4 driver line or the UAS-Dam-fusion line. We have successfully expressed Dam-fusion proteins in time windows ranging from 6 to ~72 h. Time windows may need to be optimized, depending on the fusion protein and cell type being profiled: 16–24 h works well for many tissues and GAL4 drivers[1,7]. Another option for temporal control of the GAL4 system is the GeneSwitch, which involves feeding *Drosophila* RU486 to activate gene expression[31].

Controls. To control for nonspecific methylation of DNA by the Dam methylase, DamID and TaDa experiments require that a Dam-only experiment be performed in parallel with the Dam-fusion experiment (the Dam-only fly line is available upon request to A.H.B.)[2,7]. The decision to use the whole organism or to dissect the tissue depends on the specificity of the driver and how many cells are being profiled relative to the total number of cells in the sample, although some degree of dissection is generally recommended when working with larvae or adult flies.

Sequencing. Libraries can be prepared using any NGS kit, by modifying the amplification step as appropriate to prevent concatemer formation. The software pipeline is optimized to work with single-end reads (we recommend a 50-nt read length), although BAM files generated from paired-end reads are compatible. For the *Drosophila* genome, samples should be sequenced so as to obtain 20M mapable reads.

Data analysis. The damidseq_pipeline software should generate accurate binding profiles with default settings for almost all Dam-fusion proteins analyzed. An exception to the default settings may arise when profiling proteins that are completely sequestered from open chromatin—in these circumstances, using the optional readcount (RPM) normalization procedure may generate more appropriate binding profiles.

MATERIALS

REAGENTS

- Dam-fusion (user-generated) and Dam-only (available from A.H.B. on request) flies crossed to an appropriate GAL4 driver line (available from *Drosophila* stock centers, e.g., Bloomington Drosophila Stock Center, http://flystocks.bio.indiana.edu/)
- Autoclaved dH_2O
- EDTA (Sigma-Aldrich, cat. no. E6758)
- QIAamp DNA Micro Kit (Qiagen, cat. no. 56304)
- Ethanol (EtOH; 100%; VWR, cat. no. 20821.330)
- DpnI and CutSmart buffer (NEB, cat. no. R0176S)
- PCR Purification Kit (Qiagen, cat. no. 28104)
- dsADR adaptors (0.05 µmol, desalted, from, e.g., Sigma); see **Table 1** for oligo sequence and Reagent Setup for preparation
- NGS Primer Mix; see **Table 1** for oligo sequences and Reagent Setup for preparation
- Tris, 10 mM, pH 8.0 (Sigma-Aldrich, cat. no. T2694)
- Sequencing (NGS) adaptors (0.05 µmol, DST purity from, e.g., Sigma); see **Supplementary Table 1** for oligo sequence
- DpnII and DpnII buffer (NEB, cat. no. R0543S)
- Advantage 2 cDNA polymerase (Clontech, cat. no. 639201); MyTaq HS DNA polymerase (Bioline, cat. no. BIO-21112) can be used as an alternative polymerase (S.S. de Vries and B. van Steensel, personal communication)

▲ **CRITICAL** A polymerase with hotstart is required to prevent primer-dimer formation during PCR.

- AlwI (NEB, cat. no. R0513S)
- RNase A (DNase free; Roche, cat. no. 11119915001)
- Qubit assay tubes (Invitrogen, cat. no. Q32856)
- Qubit dsDNA HS assay kit (Invitrogen, cat. no. Q32851)
- Reagents for TapeStation: Genomic DNA ScreenTape (Agilent, cat. no. 5067-5365) and reagents (Agilent, cat. no. 5067-5366)
- Reagents for Bioanalyzer: DNA 1000 Kit (Agilent, cat. no. 5067-1504)
- Agencourt AMPure XP Beads (Beckman Coulter, cat. no. A63880) (Seramag SpeedBeads, 3 EDAC/PA5 (Fisher Scientific, cat. no. 12326433), prepared using the method of Rohland and Reich[32] with 20% (wt/vol) PEG-8000 are a cost-effective alternative to AMPure XP beads)
- Quick Ligation Kit (NEB, cat. no. M2200S)
- T4 DNA ligase (400,000 U/ml) and 10× buffer (NEB, cat. no. M0202S)
- T4 DNA polymerase (NEB, cat. no. M0203S)
- Klenow fragment (NEB, cat. no. M0210S)
- Klenow 3′–5′ exo-enzyme (NEB, cat. no. M0541S)
- T4 polynucleotide kinase (NEB, cat. no. M0201S)
- NEBNext High-Fidelity 2× PCR Master Mix (NEB, cat. no. M0541S)
- dNTPs (NEB, cat. no. N0446S)

TABLE 1 | Oligos required for TaDa (DST purity).

Oligo name	Sequence (5′–3′)	Purpose
AdRt oligo (unmodified)	CTAATACGACTCACTATAGGGCAG CGTGGTCGCGGCCGAGGA	DamID adaptor (top strand) for amplification
AdRb oligo (unmodified)	TCCTCGGCCG	DamID adaptor (bottom strand) for amplification
DamID_PCR oligo (unmodified)	GGTCGCGGCCGAGGATC	Primer for amplifying dsAdR adaptor-ligated DNA
NGS_PCR1 (* = phosphorothioate linkage)	AATGATACGGCGACCACCGA*G	Primer for amplifying sequencing of adaptor-ligated DNA. Phosphorothioate linkages decrease degradation of the final library
NGS_PCR2 (* = phosphorothioate linkage)	CAAGCAGAAGACGGCATACGA*G	Primer for amplifying sequencing of adaptor-ligated DNA. Phosphorothioate linkages decrease degradation of the final library

EQUIPMENT

- Temperature-controlled metal heat block capable of heating to 95 °C
- Benchtop centrifuge capable of spinning at 20,000*g*
- PCR machine (no specific manufacturer or requirements)
- Sonicator (e.g., Diagenode Bioruptor, cat. no. B01020001)
- Magnetic rack (e.g., DynaMag-96 side magnet; Invitrogen, cat. no. 12331D)
- DNA fluorometer (e.g., Qiagen Qubit 3.0, cat. no. Q33216)
- DNA analyzer for sizing and quality control of DNA samples (e.g., Agilent 2200 TapeStation, cat. no. G2965AA)
- P1000 pipette
- Pestle

REAGENT SETUP

10 mM Tris, pH 8.0 To make 10 mM Tris, pH 8.0, take 1 ml of 1M solution and make up to 100 ml with dH_2O.

Generation of dsAdR adaptors See **Table 1** for adaptor sequences. Mix together 50 µl of AdRt (100 µM) and 50 µl of AdRb (100 µM) in a 1.5-ml tube, and incubate the mixture in a removable metal heating block at 95 °C for 2 min. Remove the heating block and allow the mixture to cool to room temperature (RT; 22 °C); this should take >45 min—slow cooling allows efficient annealing of the oligos). Annealed adaptors can be stored at −20 °C for >6 months.

Generation of NGS adaptors See **Supplementary Table 1** for adaptor sequences. Resuspend adaptor oligos at 100 µM in 10 mM Tris, pH 8.0 + 50 mM NaCl. Mix 50 µl of Universal adaptor with 50 µl of relevant index adaptor in a 1.5-ml tube and incubate the mixture in a removable metal heating block at 95 °C for 2 min. Remove the heating block and allow the mixture to cool to RT (should take >45 min—slow cooling allows efficient annealing of the oligos). Annealed oligos can be stored at −20 °C for >6 months.

NGS Primer Mix See **Table 1** for primer sequences. Mix 50 µl of NGS_PCR1 (50 µM) and 50 µl of NGS_PCR2 (50 µM). This mixture can be stored at −20 °C for >6 months.

10 mM dNTPs Mix 25 µl each of 100 mM dNTPs (dATP, dTTP, dCTP and dTTP) with 150 µl of dH_2O. The dNTPs can be stored at −20 °C for >6 months.

RNase A Dilute DNase-free RNase A to 12.5 µg/ml.

Resuspension buffer Mix 10 mM Tris, pH 8.0, with 0.1 mM EDTA.

Adaptor ligation buffer Prepare the buffer as shown below. The buffers may be stored as aliquots at −20 °C for up to 1 year (multiple freeze/thaw cycles of buffer aliquots should be avoided if possible).

Component	Volume (×1, µl)	Master mix (×100, µl)
10× T4 DNA ligase buffer	2	200
0.8 µl dsAdR	0.8	80
dH_2O	1.2	120

TaDa DpnII digestion buffer Prepare the buffer as shown below. The buffers may be stored as aliquots at −20 °C for up to 1 year (multiple freeze/thaw cycles of buffer aliquots should be avoided if possible).

Component	Volume (×1, µl)	Master mix (×100, µl)
10× DpnII buffer	4	400
dH_2O	15	1,500

DamID PCR buffer Prepare the buffer as shown below. The buffers may be stored as aliquots at −20 °C for up to 1 year (multiple freeze/thaw cycles of buffer aliquots should be avoided if possible).

Component	Volume (×1, µl)	Master mix (×50, µl)
10× cDNA PCR buffer	16	800
DamID_PCR primer (50 µM)	2.5	125
10 mM dNTPs	3.2	160
dH_2O	96.3	4,815

End-repair buffer Prepare the buffer as shown below. The buffers may be stored as aliquots at −20 °C for up to 1 year (multiple freeze/thaw cycles of buffer aliquots should be avoided if possible).

Component	Volume (×1, µl)	Master mix (×50, µl)
10× T4 DNA ligase buffer	3.0	150
10 mM dNTPs	1.2	60
dH_2O	3.3	165

End-repair enzyme mix Prepare the enzyme mix as shown below. This enzyme mix may be stored at −20 °C for up to 1 year; it does not need to be stored as aliquots because the enzyme mix will not freeze when stored at −20 °C.

Component	Volume (×1, µl)	Master mix (×50, µl)
T4 DNA polymerase (3 U/µl)	1.14	56.82
Klenow fragment (5 U/µl)	0.23	11.36
T4 polynucleotide kinase (10 U/µl)	1.14	56.82

PROCEDURE

Dissection of tissue (if required) ● TIMING ~30 min per condition

1| Induce GAL4 activation of the TaDa cassette 16–24 h before collection.

▲ CRITICAL STEP A Dam-only control, driven by the same experimental conditions as the Dam-fusion protein, must always be processed in parallel.

2| Dissect tissue in PBS. Collect enough material to include at least 15,000 labeled cells; include more labeled cells if expression of the fusion protein was induced for <16 h.

3| When enough cells have been collected, remove excess PBS and store the samples until required.

■ PAUSE POINT Samples can be stored at –20 or –80 °C until required.

Extraction of genomic DNA (using the QIAamp DNA Micro Kit) ● TIMING 1.5–2 h

4| This step can be performed using option A or B, depending on whether whole organisms or dissected tissues are used.

▲ CRITICAL STEP From this point until the completion of DpnI digestion, do not vortex or pipette the mixture vigorously, as this can result in shearing of the DNA. Mix samples by either very gentle flicking or gentle pipetting using a P1000.

(A) Whole embryos or other tissues requiring mechanical disruption (e.g., adult heads) ● TIMING 15 min

(i) Take samples from the freezer and add 175 µl of 1× PBS to the Eppendorf tube containing the sample. If you are processing samples containing gut tissue, we recommend the addition of EDTA in order to prevent fragmentation of genomic DNA by nucleases.
(ii) Add 20 µl of RNase (12.5 µg/ml) and mix it by pipetting.
(iii) Use a sterile pestle (washed in 100% ethanol) to homogenize the sample in PBS.
(iv) Add 20 µl of proteinase K (QIAamp DNA Micro Kit) and mix it gently; leave it for 1 min at RT.
(v) Add 200 µl of buffer AL (QIAamp DNA Micro Kit) and mix it gently; incubate the mixture at 56 °C for 10 min.
(vi) Cool the mixture to RT, add 200 µl of 100% ethanol and mix it gently.

(B) Small volumes (<10 µl total volume) of dissected tissue ● TIMING 1–6 h

(i) Take cut larvae or dissected tissue from the freezer. If you are processing samples containing gut tissue, we recommend the addition of EDTA in order to prevent fragmentation of genomic DNA by nucleases.
(ii) Add 180 µl of ATL buffer (QIAamp DNA Micro Kit).
(iii) Add 20 µl of proteinase K (QIAamp DNA Micro Kit) and mix it gently.
(iv) Incubate the mixture at 56 °C on a heat block until digested; flick gently to mix the contents occasionally. ▲ **CRITICAL STEP** Digestion will typically take 1 h for most tissues, but it may take longer if larger volumes of material are used.
(v) Cool the mixture to RT and add 20 µl of RNase (12.5 µg/ml).
(vi) Incubate the mixture at RT for 2 min.
(vii) Mix 200 µl of buffer AL (QIAamp DNA Micro Kit) and 200 µl of 100% EtOH together (per sample) in a separate tube; mix them well by vortexing.
(viii) Add 400 µl of buffer AL/EtOH mix to each sample; mix well by inversion.

5| Add all of the solution (from Step 4A(vi) or 4B(viii)) to a spin column (QIAamp DNA Micro Kit).

6| Spin it at >6,000*g* at room temperature for 1 min, and discard the flow-through and the collecting tube.

7| Add 500 µl of AW1 solution (QIAamp DNA Micro Kit) and spin the mixture at >6,000*g* for 1 min; discard the flow-through and the collecting tube.

8| Add 500 µl of AW2 (QIAamp DNA Micro Kit) solution and spin the mixture at >6,000*g* for 1 min; discard the flow-through and the collecting tube.

9| Spin the mixture at 20,000*g* for 3 min to dry the column.

10| Transfer the spin column to a new 1.5-ml tube, add 50 µl of AE buffer (QIAamp DNA Micro Kit) and leave it at RT for at least 30 min. Spin it at >6,000*g* for 1 min and keep the flow-through, which contains the genomic DNA.

11| Run 1 µl of genomic DNA on a 0.8% agarose gel to check the quality. The genomic DNA should be visible as a single band at the top of the gel and not as a smear.
? TROUBLESHOOTING

Digestion of methylated DNA ● TIMING ~16 h

12| Transfer 43.5 µl of genomic DNA to a new 1.5-ml tube.

13| Add 5 µl of NEB CutSmart Buffer and 1.5 µl of DpnI enzyme, mix it very gently either by flicking or with a P1000 pipette and digest the mixture overnight (at least 12 h) at 37 °C.
▲ **CRITICAL STEP** Do not vortex the DpnI solution, as this can result in shearing of the genomic DNA.

14| Clean up the DpnI-digested DNA with the Qiagen PCR Purification Kit according to the manufacturer's instructions elute in 32 µl of H_2O.
■ **PAUSE POINT** DNA can be stored for up to 6 months at 20 °C.

Ligation of DamID adaptors to DpnI-digested DNA ● TIMING 2.5 h

15| Measure the DNA concentration using Qubit or NanoDrop fluorometer and either use 15 µl of undiluted sample (if less than a total of 750 ng) or dilute the sample with H_2O to a total of 750 ng in 15 µl for the next step (the remaining DNA can be stored at –20 °C). Depending on the starting material and the proportion of induced cells, very low to undetectable DNA concentrations at this step are completely normal.

16| Transfer 15 µl of sample to 0.2-ml PCR tubes and add 4 µl of premade adaptor ligation buffer (see Reagent Setup section).

17| Add 1 µl (400 U) of T4 DNA ligase enzyme and mix it well.

18| Incubate the ligation reaction for 2 h at 16 °C in a PCR machine, followed by 10 min at 65 °C to inactivate the ligase.

DpnII digestion of DNA ● TIMING 2 h

19| Add 19 µl of premade TaDa DpnII digestion buffer (see Reagent Setup section).

20| Add 1 µl of DpnII enzyme and mix it well.

21| Digest the mixture at 37 °C (use an incubator or a PCR machine) for at least 2 h (a longer digestion is acceptable).

PCR amplification ● TIMING 2.5 h

22| Add 118 µl of premade DamID PCR buffer (see Reagent Setup section) to the DpnII-digested DNA.

23| Add 2 µl of Advantage 2 cDNA polymerase enzyme and mix well.

24| Split the reaction into 4 × 40 µl reactions in 0.2-ml PCR tubes.
▲ **CRITICAL STEP** For an efficient PCR amplification, the PCR volume should be kept low (≤40 µl). Larger reaction volumes can result in inefficient amplification of large DNA fragments.

25| Perform PCR using the following program:

Cycle number	Denature	Anneal	Extend
0			68 °C, 10 min
1	94 °C, 30 s	65 °C, 5 min	68 °C, 15 min
2–4	94 °C, 30 s	65 °C, 1 min	68 °C, 10 min
5–21	94 °C, 30 s	65 °C, 1 min	68 °C, 2 min
22			68 °C, 5 min

■ **PAUSE POINT** PCR-amplified DNA can be stored overnight at 4 °C or for up to 6 months at 20 °C.

Sonication and removal of DamID adaptors ● TIMING ~16 h

26| Run 3 µl of the PCR product on a 0.8% agarose gel to check the quality. A successful amplification should show a smear concentrated between 400 bp and 2 kb; some unique bands may also be present, depending on the particular binding characteristics of the Dam-fusion protein. See **Figure 3** for an example.
? TROUBLESHOOTING

27| Use the Qiagen PCR Purification Kit according to the manufacturer's instructions to purify the rest of the PCR product, and use 32 µl of H_2O to elute (leave water on the column for at least 5 min before the final spin).

28| Measure the concentration of purified DNA using a Qubit or NanoDrop fluorometer, and dilute 2 µg of the purified DNA in 90 µl of H_2O in a 1.5-ml tube.
? TROUBLESHOOTING

29| Add 10 µl of NEB CutSmart Buffer and mix it well.

30| Sonicate the DNA. Sonication time and conditions are highly dependent upon the equipment used, and they should be optimized to generate an average DNA fragment size of ~300 bp. (With a Bioruptor Plus, we use 5 min, high power, 30 s on/30 s off).

31| Check the sonicated DNA on an Agilent TapeStation or a Bioanalyzer. The fragments should have a mean size of ~300 bp (see **Fig. 3** for an example).

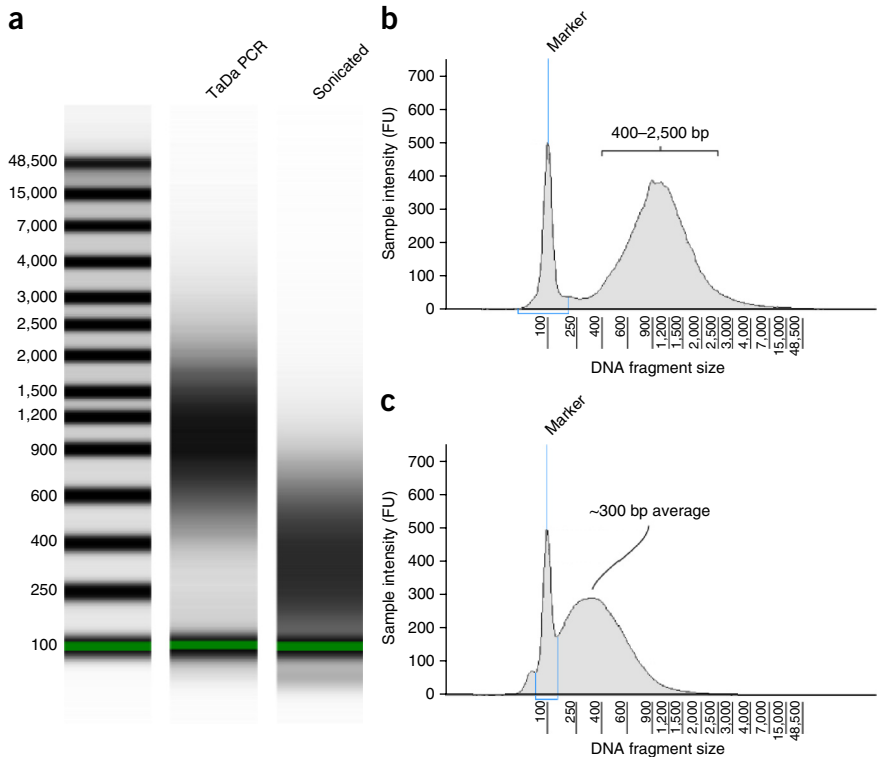

Figure 3 | Sample TapeStation plots showing the expected fragment-size distribution following TaDa. (**a**) Gel image of samples following the TaDa PCR step (Step 25) and sonication (Step 30). (**b**) Electropherogram of the TaDa PCR sample. (**c**) Electropherogram of the sample after sonication. The sample illustrated was prepared by driving the DAM-only construct in adult neurons for 24 h, with DNA isolated from whole adult heads. FU, fluorescence units.

32| Remove the DamID adaptors from the sonicated DNA by adding 1 µl of AlwI (mix well). Incubate the mixture overnight at 37 °C.
? TROUBLESHOOTING

33| Transfer 70 µl of each sample to PCR strips for library preparation.
■ **PAUSE POINT** Sonicated DNA can be stored for up to 6 months at 20 °C.

Sequencing library preparation: DNA cleanup ● TIMING 40 min

34| Add 105 µl of AMPure XP beads to 70 µl of the sample and mix well.
▲ **CRITICAL STEP** Ensure that the bead stock is fully mixed before pipetting (use vortex).

35| Incubate the mixture at RT for 10 min.

36| Place the mixture on a magnetic stand for 10 min (or until clear) and keep the tubes on the magnetic stand for the duration of Steps 37–39.

37| Remove and discard the supernatant.
▲ **CRITICAL STEP** Do not touch the bead pellet with the pipette tip.

38| Wash the beads twice in 190 µl of 80% EtOH (leave beads on the magnet; 30-s wash time).

39| Remove as much liquid as possible, and leave the beads for 5 min to air-dry.

40| Remove the PCR strip from the magnetic stand and resuspend the beads in 25 µl of resuspension buffer.

41| Incubate the mixture for 2 min at RT.

42| Place the mixture on a magnetic stand for 5 min (or until clear).

43| Put 22.5 µl of the supernatant into a new, clean tube for the next step.
■ **PAUSE POINT** Purified DNA can be stored for up to 6 months at 20 °C.

Sequencing library preparation: adjust concentrations ● TIMING 20 min

44| Use 1 µl of supernatant to measure the sample concentration on a Qubit or similar fluorometer.

45| Dilute the remaining sample to (no more than) 500 ng of DNA in 20 µl of resuspension buffer (remainder can be discarded).

Sequencing library preparation: end repair ● TIMING 55 min

46| Add 7.5 µl of end-repair buffer to the diluted sample from Step 45.

47| Add 2.5 µl of end-repair enzyme mix and mix it well.

48| Incubate the mixture for 30 min at 30 °C.

49| Heat-inactivate the enzymes for 20 min at 75 °C.
■ **PAUSE POINT** End-repaired DNA can be stored for up to 6 months at –20 °C.

Sequencing library preparation: adenylation of 3′ ends ● TIMING 35 min
50| Add 0.75 µl of Klenow 3′–5′ exo-enzyme to the sample from Step 49.

51| Incubate the mixture at 37 °C for 30 min.

Sequencing library preparation: sequencing adaptor ligation ● TIMING 35 min
52| Add 2.5 µl of the relevant sequencing adaptor to the sample from Step 51.
▲ **CRITICAL STEP** If multiplexing four or fewer libraries, selecting adaptors with barcodes that are too similar may result in a reduced number of reads passing the filter. In this case, the preferred indexes to use (in order) are 4, 7, 6 and 8.

53| Add 2.5 µl of NEB quick ligase enzyme.

54| Incubate the mixture at 30 °C for 10 min.

55| Add 5 µl of 0.5 M EDTA to stop the ligation reaction.
■ **PAUSE POINT** Ligated DNA can be stored for up to 6 months at –20 °C.

Sequencing library preparation: DNA cleanup ● TIMING 60 min
▲ **CRITICAL** Two rounds of bead cleanup (Steps 56–65 and 66–75) are required to ensure the complete removal of adaptor dimers.
56| Add 41 µl of AMPure beads to the sample from Step 55 and mix well.

57| Incubate the mixture at RT for 10 min.

58| Place the mixture on a magnetic stand for 5 min (or until clear).

59| Remove and discard 80 µl of supernatant.

60| Wash the beads twice in 200 µl of 80% EtOH (leave the beads on the magnet; 30-s wash time).

61| Leave the beads for 5 min to air-dry.

62| Remove the PCR strip from the magnetic stand and resuspend the beads in 52.5 µl of resuspension buffer.

63| Incubate the mixture for 2 min at RT.

64| Place the mixture on a magnetic stand for 5 min (or until clear).

65| Transfer 50 µl of the supernatant to a new, clean tube for the next round of cleanup (Steps 66–75).
■ **PAUSE POINT** Purified DNA can be stored for up to 6 months at –20 °C.

66| Add 50 µl of AMPure beads to the supernatant from Step 65.

67| Incubate the mixture at RT for 10 min.

68| Place the mixture on a magnetic stand for 5 min (or until clear).

69| Remove and discard 95 µl of supernatant.

70| Wash the beads twice in 200 µl of 80% EtOH (leave the beads on the magnet; 30-s wash time).

71| Leave the beads for 5 min to air-dry.

72| Remove the PCR strip from the magnetic stand and resuspend the beads in 22.5 µl of resuspension buffer.

73| Incubate the mixture for 2 min at RT.

74| Place the mixture on a magnetic stand for 5 min (or until clear).

75| Transfer 20 µl of the supernatant to a new, clean tube for the next step.
■ **PAUSE POINT** Purified DNA can be stored for up to 6 months at 20 °C.

Sequencing library preparation: enriching DNA fragments ● TIMING 25–40 min
76| Add 5 µl of PCR Primer Mix to the supernatant from Step 75.

77| Add 25 µl of PCR Master Mix.

78| Perform PCR as shown below:

Cycle number	Denature	Anneal	Extend
0	98 °C, 30 s		
1–6	98 °C, 10 s	60 °C, 30 s	72 °C, 30 s
7			72 °C, 5 min

▲ **CRITICAL STEP** More than six cycles may need to be performed if the amount of starting material is less than recommended. However, it is important not to use too many cycles, as this can result in adaptor concatemers.

Sequencing library preparation: DNA cleanup ● TIMING 30 min
79| Add 50 µl of AMPure beads to the 50-µl PCR sample from Step 78.

80| Incubate the mixture at RT for 10 min.

81| Place the mixture on a magnetic stand for 5 min (or until clear).

82| Remove and discard 95 µl of supernatant.

83| Wash the beads twice in 200 µl of 80% EtOH (leave the beads on the magnet; 30-s wash time).

84| Leave the beads for 5 min to air-dry.

85| Remove the PCR strip from the magnetic stand and resuspend the beads in 32.5 µl of resuspension buffer.

86| Incubate the mixture for 2 min at RT.

87| Place the mixture on the magnetic stand for 5 min (or until clear).

88| Transfer 30 µl of the supernatant to a new, clean tube for the next step.

Sequencing library quality control ● TIMING 60 min
89| Check the DNA from Step 88 on an Agilent TapeStation or a Bioanalyzer (see ANTICIPATED RESULTS) and note the average fragment size.
? TROUBLESHOOTING

90| Measure the DNA concentration with a Qubit fluorometer (if starting with 500 ng of material, the final concentration should be between 15 and 30 ng/µl).

91| Pool the libraries to a final recommended concentration for Illumina sequencing (DNA molarity = ~(1,500/(fragment size in bp)) × (concentration in ng/μl)). Ensure that all libraries have the same final concentration when pooled.
? TROUBLESHOOTING

92| Sequence single-ended 50-nt (SE50) reads on an Illumina sequencer. For *Drosophila* samples, aim for at least 20 million reads per library.
▲ **CRITICAL STEP** Other organisms may require a considerably higher number of reads per library, depending upon the genome size.

93| Process Illumina sequencing data as outlined in **Box 1**.

? TROUBLESHOOTING
Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
11	Insufficient genomic DNA	Insufficient starting material	Use more starting material (Step 2)
		Sample is not homogenized enough	Although not optimal, as little as 1 μg of genomic DNA can be used
			Perform more vigorous homogenization (Step 4)
26	Low yield after PCR amplification	Additional DNA purification step required to ensure optimal PCR conditions	Purify the DNA following DpnII digestion (after Step 21) with 60 μl of Seramag beads (see Steps 79–89) to increase PCR yield. Following elution, prepare a single 50-μl PCR reaction as follows: 30 μl purified DNA, 5 μl 10× Advantage 2 PCR Buffer, 2.5 μl DamID_PCR primer, 1 μl of 10 mM dNTPs, 10 μl H_2O and 1.5 μl of Advantage 2 cDNA polymerase. Run the PCR as described in Step 25 and at Step 26 run 1 μl of the PCR product on a 0.8% gel to check the quality. Proceed with the protocol as normal from Step 27
28	Insufficient PCR product	Small number of cells expressing the Dam-fusion protein relative to the rest of the tissue	Increase the amount of starting material at Step 2 or the length of GAL4 induction time at Step 1
32	DNA fragment sizes are too large	Insufficient sonication	Re-optimize the sonication time (Step 30)
89	Secondary peak in sequencing library	Exhaustion of PCR in the sequence library preparation, resulting in adaptor concatemers	Reduce amplification cycles (Step 78) or reduce input DNA quantity (Step 45). It is not advised to perform fewer than six cycles
	No DNA after sequencing library preparation	Problem with a reagent	Use fresh 80% ethanol when washing bead preps
			Use new AmpureXP beads
			Use fresh enzymes for end repair, tailing, ligation and PCR
91	Low numbers of reads, or a low percentage of reads mapped back to the genome	Contamination with foreign DNA	Use fresh AlwI enzyme (Step 32) and check that the correct buffer is being used (Step 29); failure to remove the original adaptors will cause problems with clustering on the Illumina sequencer and greatly reduce the number of reads passing the filter
		Failure to remove initial adaptors with AlwI	
		Adaptor dimers or adaptor concatemers present in sequencing library	Adaptor dimers or concatemers will be visible at Step 87. Concatemers are prevented by reducing PCR cycles or reducing template DNA (see above). Dimers should be removed by bead purification (Steps 54–73). Repeat, using fresh beads, and check the pipette accuracy

● **TIMING**
A schematic of the PROCEDURE and time line can be found in **Figure 2**.
Steps 1–3, pre-day 1: ~30 min per condition
Steps 4–14, day 1
Steps 15–25, day 2
Steps 26–33, day 3

Steps 34–75, day 4
Steps 76–91, day 5
Box 1, processing of sequencing data: 60 min per sample

ANTICIPATED RESULTS

Amplification of Dam-methylated DNA (Steps 1–25)

The PCR amplification should produce DNA fragments ranging from ~400 to 2,500 bp (**Fig. 3a**,**b**). The total yield may range from 2 to 20 µg, depending on the number of labeled cells and the factor being profiled.

Sonication and removal of adaptors (Steps 26–33)

Sonication of the amplified DNA should provide fragments with an average size of 300 bp (±50 bp). See **Figure 3a**,**c** for an example. DamID adaptors are subsequently removed with AlwI digestion.

Sequencing library preparation (Steps 34–93)

Final library concentration should be in excess of 50 nM, with fragments spread around a mean size of 450 bp (note that the apparent size on the TapeStation will be larger, typically 500–550 bp).

We assess overall quality through the percentage of mapped reads, which ideally should be higher than 90% in *Drosophila*. Lower percentages of mapped reads may indicate primer dimers or library contamination. The final $\log_2$ ratio profile from a DamID-seq experiment should show strong signal for bound regions with low background (see **Fig. 4** for an example profile using RNA Pol II). If you are profiling a transcription factor, peaks should be detectable through software (**Box 1**); if you are profiling transcription via RNA Pol II, genes with a false discovery rate (FDR) of <0.01 should be detectable following software analysis (**Box 1**).

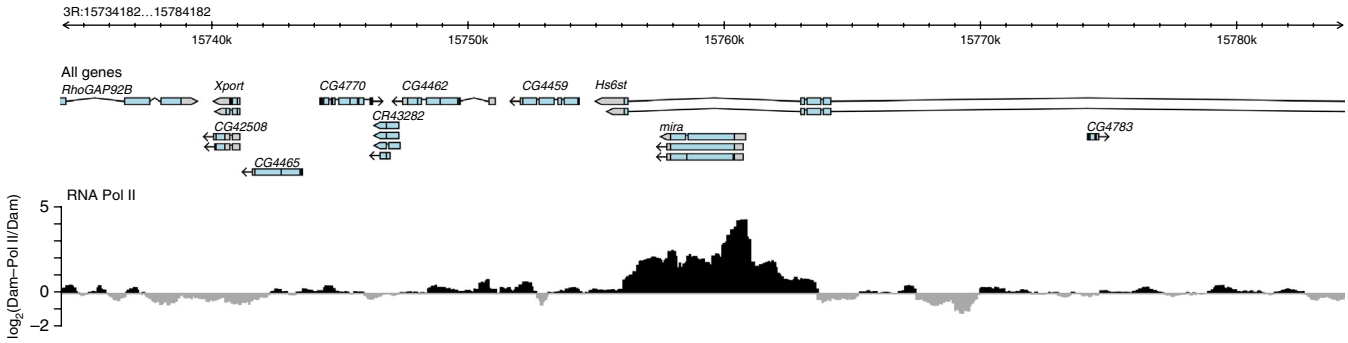

Figure 4 | Example $\log_2$ ratio profile of RNA Pol II binding in *Drosophila* neural stem cells. DNA was isolated from 50 *D. melanogaster* larvae for both Dam–Pol II and Dam-only samples; larvae were torn in half, and the anterior halves (which contained the brain) were retained for DNA isolation and processing. Expression of the TaDa proteins was driven using a neural-stem-cell-specific driver (worniu-GAL4) in the presence of tub-GAL80^{ts} and induced for 16 h at 29 °C from 72 h after larval hatching. The gene *miranda* (*mira*) illustrated here is required for the correct segregation of cell-fate determinants during the division of *Drosophila* neural stem cells.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS This work was funded by a Wellcome Trust Senior Investigator Award (103792), Wellcome Trust Programme Grant (092545) and BBSRC Project Grant (BB/L00786X/1) to A.H.B. A.H.B. acknowledges core funding to the Gurdon Institute from the Wellcome Trust (092096) and CRUK (C6946/A14492).

AUTHOR CONTRIBUTIONS O.J.M., T.D.S., S.W.C. and A.H.B. designed the research and analyzed the data; T.D.S., S.W.C. and O.J.M. developed the protocols; and O.J.M., T.D.S., S.W.C. and A.H.B. wrote the paper.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

1. Southall, T.D. *et al.* Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. *Dev. Cell.* **26**, 101–112 (2013).
2. van Steensel, B. & Henikoff, S. Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).
3. Vogel, M.J., Peric-Hupkes, D. & van Steensel, B. Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478 (2007).
4. Brand, A.H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
5. Luukkonen, B.G., Tan, W. & Schwartz, S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J. Virol.* **69**, 4086–4094 (1995).
6. Van Blokland, H.J.M., Hoeksema, F., Siep, M., Otte, A.P. & Verhees, J.A. Methods to create a stringent selection system for mammalian cell lines. *Cytotechnology* **63**, 371–384 (2011).
7. Marshall, O.J. & Brand, A.H. damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **31**, 3371–3373 (2015).
8. Filion, G.J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).

9. Woolcock, K.J., Gaidatzis, D., Punga, T. & Bühler, M. Dicer associates with chromatin to repress genome activity in *Schizosaccharomyces pombe*. *Nat. Struct. Mol. Biol.* **18**, 94–99 (2011).
10. Germann, S., Juul-Jensen, T., Letarnec, B. & Gaudin, V. DamID, a new tool for studying plant chromatin profiling *in vivo*, and its use to identify putative LHP1 target loci. *Plant J.* **48**, 153–163 (2006).
11. Schuster, E. *et al.* DamID in *C. elegans* reveals longevity-associated targets of DAF-16/FoxO. *Mol. Syst. Biol.* **6**, 399 (2010).
12. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
13. Vogel, M. *et al.* Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* **16**, 1493–1504 (2006).
14. Bryant, Z. *et al.* Characterization of differentially expressed genes in purified *Drosophila* follicle cells: toward a general strategy for cell type-specific developmental analysis. *Proc. Natl. Acad. Sci. USA* **96**, 5559–5564 (1999).
15. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).
16. Bonn, S. *et al.* Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat. Protoc.* **7**, 978–994 (2012).
17. Deal, R.B. & Henikoff, S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell* **18**, 1030–1040 (2010).
18. Henry, G.L., Davis, F.P., Picard, S. & Eddy, S.R. Cell type-specific genomics of *Drosophila* neurons. *Nucleic Acids Res.* **40**, 9691–9704 (2012).
19. Steiner, F.A., Talbert, P.B., Kasinathan, S., Deal, R.B. & Henikoff, S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.* **22**, 766–777 (2012).
20. Schauer, T. *et al.* CAST-ChIP maps cell-type-specific chromatin states in the *Drosophila* central nervous system. *Cell Rep.* **5**, 271–282 (2013).
21. Otsuki, L., Cheetham, S.W. & Brand, A.H. Freedom of expression: cell-type-specific gene profiling. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 429–443 (2014).
22. Roy, P.J., Stuart, J.M., Lund, J. & Kim, S.K. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**, 975–979 (2002).
23. Thomas, A. *et al.* A versatile method for cell-specific profiling of translated mRNAs in *Drosophila*. *PLoS ONE* **7**, e40276 (2012).
24. Miller, M.R., Robinson, K.J., Cleary, M.D. & Doe, C.Q. TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat. Methods* **6**, 439–441 (2009).
25. Modrich, P. & Lahue, R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* **65**, 101–133 (1996).
26. Baena-Lopez, L.A., Alexandre, C., Mitchell, A., Pasakarnis, L. & Vincent, J.-P. Accelerated homologous recombination and subsequent genome modification in *Drosophila*. *Development* **140**, 4818–4825 (2013).
27. Venken, K.J.T. *et al.* MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nat. Methods* **8**, 737–743 (2011).
28. Luan, H., Peabody, N.C., Vinson, C. & White, B.H. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron* **52**, 425–436 (2006).
29. Matsumoto, K., Toh-e, A. & Oshima, Y. Genetic control of galactokinase synthesis in *Saccharomyces cerevisiae*: evidence for constitutive expression of the positive regulatory gene gal4. *J. Bacteriol.* **134**, 446–457 (1978).
30. McGuire, S.E., Le, P.T., Osborn, A.J., Matsumoto, K. & Davis, R.L. Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765–1768 (2003).
31. Osterwalder, T., Yoon, K.S., White, B.H. & Keshishian, H.A Conditional tissue-specific transgene expression system using inducible GAL4. *Proc. Natl. Acad. Sci. USA* **98**, 12596–12601 (2001).
32. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).