

**Editorial summary:** Targeted DamID (TaDa) extends DamID to enable cell-type specific profiling of genome-wide protein binding. Transcription factor binding sites, RNA Pol II occupancy and chromatin states can be studied to provide insights into cell fate specification.

**Proposed Tweet:** Cell-type specific profiling of protein-DNA interactions with TaDa [LINK] #chromatin #transcription

### **Ontology**

Biological sciences / Biological techniques / Gene expression analysis

Biological sciences / Genetics / Genomics / Transcriptomics

Biological sciences / Molecular biology / Transcription

Biological sciences / Molecular biology / Chromatin

**Keywords:** DamID, Targeted DamID, cell type specific, transcriptional profiling, chromatin profiling, RNA polymerase II, RNA Pol II, RNAPII, genome wide, next generation sequencing

### **Supporting primary publications:**

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**, 101–12 (2013). doi: 10.1016/j.devcel.2013.05.020

Marshall, O. J. & Brand, A. H. damidseq\_pipeline: an automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **31**, 3371–3 (2015). doi: 10.1093/bioinformatics/btv386

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

Owen J. Marshall\*, Tony D. Southall\*# and Andrea H. Brand

The Gurdon Institute and Department of Physiology, Development and Neuroscience, University of Cambridge, Tennis Court Road, Cambridge, United Kingdom CB2 1QN

\* These authors contributed equally to this work.

# Current address: Department of Life Sciences, Imperial College London, Sir Ernst Chain Building, South Kensington Campus, London SW7 2AZ, UK

Corresponding author:

Andrea H. Brand; a.brand@gurdon.cam.ac.uk

Telephone: 00-44-1223-334141; FAX: 00-44-1223-334089

## Summary

The ability to profile transcription and chromatin binding in a cell type-specific manner is a powerful approach for 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 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, avoiding toxicity and potential artefacts from over-expression. 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 to other methods for cell-type specific profiling, the technique requires no cell-sorting, crosslinking 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, 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 days.

## Introduction

Understanding 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 to investigate the mechanisms that specify cell fate and cell properties.

### Development of Targeted DamID

We recently developed “TaDa” (Targeted DamID) to profile genome-wide, cell-type-specific protein binding *in vivo* without cell isolation<sup>1</sup>. 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)<sup>2,3</sup>, an *in vivo* chromatin profiling technique. For DamID, *E. 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. Expression of Dam at high levels is toxic, however and as a result DamID requires the Dam-fusion protein to be expressed at extremely low levels<sup>2</sup>. 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<sup>4</sup>, results in excessively high levels of the methylase and toxicity<sup>1</sup>. 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 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 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.<sup>1,5,6</sup> 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 only cuts at adenine-methylated GATC sites. The digested DNA is column purified to exclude un-cut, 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 only cuts non-methylated GATC sites (steps 19-21). The resulting DNA is then used as template for a PCR reaction to amplify the methylated fragments (steps 22-25). The product of the PCR reaction will thus be highly enriched for bound GATC fragments. Although based on the original DamID technique<sup>3</sup>, 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 NGS (steps 26-30), as some DNA fragments generated by the DamID procedure are too large for sequencing. In order to increase the mapping specificity of reads and overcome problems of low-diversity, DamID adaptors are removed with AlwI (steps 31-33), which removes the DamID adaptors and 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 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 are reduced. Following 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 polymerase II) 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 *D. melanogaster*, including embryonic and larval neural stem cells, and larval and adult neurons (<sup>1,7</sup>; Brand lab, unpublished). To assess transcription, we fused Dam to RNA polymerase II (Pol II), enabling the identification of transcribed

genes in a cell type-specific manner<sup>1</sup>. 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, for example by simultaneously profiling with Dam-Pol II whilst knocking down a gene using RNAi (P. Speder and A.H. Brand, unpublished). 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. Brand, in preparation). Chromatin states can be broadly classified into five different types, which can be identified by profiling five different proteins with DamID<sup>8</sup>. 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. Marshall and A.H. Brand, in preparation). 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<sup>9</sup>, plants<sup>10</sup>, *C. elegans*<sup>11</sup>, and human tissue culture cells<sup>12,13</sup>. 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. Cheatham, W. Gruhn, A. Surani and A.H. Brand, unpublished).

### **Comparisons with other methods**

A number of techniques currently exist for performing cell-type specific profiling, which take one of two approaches. The first approach focuses on specifically labelling and isolating the subset of cells of interest before undertaking conventional ChIP profiling. Examples include BITS-ChIP, where cells<sup>14</sup> or fixed nuclei<sup>15,16</sup> are isolated by fluorescent activated cell sorting (FACS), and the INTACT method, where nuclei are isolated by biotin labelling and pulldown<sup>17-19</sup>. 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<sup>6</sup> cells for ChIP), can be technically challenging, and can potentially incorporate unlabelled 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<sup>20</sup> 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 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<sup>21</sup>, via the targeted expression of a tagged RNA-binding protein<sup>22</sup>, a ribosomal protein<sup>23</sup>, or an RNA modifying enzyme<sup>24</sup>. These methods often require large amounts of starting material and immunoprecipitation.

In 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 artefacts that may arise from fixation, and does so with very low levels of protein translation (so low that TaDa fusion proteins are undetectable via western blot<sup>1</sup>), thus minimising the impact of protein over-expression. 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, thus a 1:2000 ratio of methylated DNA to total DNA), using only 100 heads and thus ~10,000 cells in total (T. D. Southall and A. H. Brand, unpublished).

### 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 ~190bp). Therefore it cannot always provide the same resolution as ChIP-seq experiments. TaDa profiling of Pol II occupancy does not distinguish between all alternatively spliced transcripts and 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 recognise these modifications.

### Experimental design

**Dam-fusion protein expression vector.** In order to obtain inducible expression coupled with low-level translation, DNA-binding proteins of interest should be cloned into the pUAST-attB-LT3-NDam vector<sup>1</sup> (see Supplementary Fig. 1; full vector sequence available from Genbank acc. no KU728166; vector available upon request to AHB). 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 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 following 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 due to the role of adenine methylation in mismatch repair in *E. coli*<sup>25</sup>. Bacteria carrying mutated Dam appear to have a growth advantage, therefore we tend to select the smaller colonies following transformation of the plasmid.

**Cell-type specific expression.** To profile protein-DNA interactions in cells of interest the choice of GAL4 driver line is important. TaDa is highly sensitive and, therefore, it is important that the GAL4 line is specific to the cells of interest. Drivers generated by homologous recombination<sup>26</sup> or by conversion of MiMIC lines<sup>27</sup> are usually optimal. The split-GAL4 system<sup>28</sup> 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<sup>ts</sup>, can be utilised<sup>29,30</sup>. GAL80<sup>ts</sup> inactivates GAL4 at 18°C but not at 29°C. With GAL80<sup>ts</sup> 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<sup>ts</sup> (tub-GAL80<sup>ts</sup>) 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 6hrs to ~72hrs (<sup>1,7</sup>; D. Doupé and A.H. Brand, unpublished). Time windows may need to be optimised, depending upon the fusion protein and cell-type being profiled: 16-24 hours works well for many tissues and GAL4 drivers.<sup>1,7</sup> Another option for temporal control of the GAL4 system is the GeneSwitch, which involves feeding *Drosophila* RU486 to activate gene expression.<sup>31</sup>

**Controls.** To control for non-specific methylation of DNA by the Dam methylase, DamID and TaDa experiments require a Dam-only experiment to be performed in parallel with the Dam-fusion experiment (the Dam-only fly line is available upon request to AHB).<sup>2,7</sup> 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, modifying the amplification step as appropriate to prevent concatamer formation. The software pipeline is optimised to work with single-end reads (we recommend 50nt 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 analysed. 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) normalisation procedure may generate more appropriate binding profiles.

## MATERIALS

### REAGENTS

Dam-fusion (user-generated) and Dam-only (available from AHB on request) flies crossed to an appropriate GAL4 driver line (available from *Drosophila* stock centres e.g. Bloomington)

Autoclaved dH<sub>2</sub>O

EDTA (Sigma-Aldrich, cat. no. E6758)

QIAamp DNA micro kit (Qiagen, cat. no. 56304)

Ethanol (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 µmole, DST purity 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

Resuspension buffer: 10mM Tris pH 8.0, 0.1mM EDTA

Sequencing adaptors (0.05 µmole, DST purity from, e.g. Sigma); see Supplementary Table 1 for oligo sequence and Reagent Setup for preparation.

DpnII and DpnII buffer (NEB, cat. no. R0543S)

Advantage 2 cDNA polymerase (Clontech, cat. no. 639201)

(An alternative polymerase is MyTaq HS DNA Polymerase (Bioline, cat. no. BIO-21112); S. S. de Vries and B. van Steensel, personal communication.)

▲ **CRITICAL STEP** 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) diluted to 12.5µg/ml

Qubit assay tubes (Invitrogen, cat. no. Q32856)

Qubit dsDNA HS assay kit, (Invitrogen, cat. no. Q32851)

Agilent reagents for TapeStation: Genomic DNA ScreenTape(5067-5365) and reagents (5067-5366)

Agilent reagents for Bioanalyzer:Agilent DNA 1000 Kit (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, 2012<sup>32</sup> with 20%(w/v) PEG-8000 are a cost-effective alternative to AMPure XP beads)

Quick ligase (NEB, cat. no. M2200S)

T4 DNA ligase (400,000U/ml) and 10x buffer (NEB, cat. no. M0202S)

T4 DNA polymerase (NEB, cat. no. M0203S)

Klenow Fragment (NEB, cat. no. M0210S)

Klenow 3' to 5' exo- (NEB, cat. no. M0541S)

T4 polynucleotide kinase (NEB, cat. no. M0201S)

NEBNext High-Fidelity 2X PCR Master Mix (NEB, cat. no. M0541S)

dNTPs (NEB, cat. no. N0446S)

Table 1. Oligos required for TaDa (DST purity)

Oligo name	Sequence (5' to 3')	Purpose
AdRt oligo (unmodified)	CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGA GGA	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 adaptor-ligated DNA. Phosphorothioate linkages decrease degradation of the final library.
NGS_PCR2 (* = Phosphorothioate linkage)	CAAGCAGAAGACGGCATAACGA*G	Primer for amplifying sequencing adaptor-ligated DNA. Phosphorothioate linkages decrease degradation of the final library.



## EQUIPMENT

Temperature controlled heat block capable of heating to 95°C.

Benchtop centrifuge capable of spinning at 20,000 x g.

PCR machine (no specific manufacturer or requirements)

Sonicator (for example, Diagenode Bioruptor cat. no. B01020001)

Magnetic rack (for example, DynaMag-96 side magnet (Invitrogen cat. no. 12331D))

DNA fluorometer (for example, Qiagen Qubit 3.0 cat. no. Q33216)

DNA analyser for sizing and quality control of DNA samples (for example, Agilent 2200 TapeStation cat. no. G2965AA)

## REAGENT SETUP

**Generation of dsAdR adaptors** Mix together 50 µl AdRt (100 µM) and 50 µl AdRb (100 µM) in a 1.5 ml tube and incubate in removable metal heating block at 95°C for 2 mins. Remove heating block and allow to cool to room temperature (should take > 45mins – slow cooling allows efficient annealing of the oligos). Annealed adaptors can be stored at -20°C for >6 months.

**Generation of NGS adaptors** Resuspend adaptor oligos at 100µM in TE + 50mM NaCl. Mix 50µl Universal adaptor with 50µl of relevant index adaptor in a 1.5ml tube and incubate in removable metal heating block at 95°C for 2 mins. Remove heating block and allow to cool to room temperature (should take > 45mins – slow cooling allows efficient annealing of the oligos). Annealed oligos can be stored at -20°C for >6 months.

**NGS Primer Mix** Mix 50µl NGS\_PCR1 (50µM) and 50µl NGS\_PCR2 (50µM). Can be stored at -20°C for >6 months.

**10mM dNTPs** Mix 25µl each of 100mM dNTPs (dATP, dTTP, dCTP and dTTP) with 150µl dH<sub>2</sub>O. Can be stored at -20°C for >6 months.

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

Component	Volume (x1)	Master mix x100
10x T4 DNA ligase buffer	2 µl	200 µl
0.8µL dsAdR	0.8 µl	80 µl
dH <sub>2</sub> O	1.2 µl	120 µl

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

Component	Volume (x1)	Master mix x100
10x Dpn II Buffer	4 µl	400 µl
dH <sub>2</sub> O	15 µl	1500 µl



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

Component	Volume	
	(x1)	Master mix x50
10X cDNA PCR reaction buffer	16 µl	800 µl
DamID_PCR primer (50 µM)	2.5 µl	125 µl
10mM dNTPs	3.2 µl	160 µl
dH <sub>2</sub> O	96.3 µl	4815 µl

**End Repair Buffer** Prepare as tabulated below. The buffers may be stored as aliquots at -20°C for up to one year (multiple freeze/thaw cycles of buffer aliquots should be avoided if possible).

Component	Volume (x1)	
	(x1)	Master mix x50
10x T4 DNA ligase buffer	3.0 µl	150 µl
10mM dNTPs	1.2 µl	60 µl
dH <sub>2</sub> O	3.3 µl	165 µl

**End Repair Enzyme mix** May be stored at -20°C for up to one year; it does not need to be stored as aliquots because the enzyme mix will not freeze when stored at -20°C.

Component	Volume (x1)	
	(x1)	Master mix x50
T4 DNA polymerase (3U/µL)	1.14 µl	56.82 µl
Klenow Fragment (5 U/µL)	0.23 µl	11.36 µl
T4 polynucleotide kinase (10 U/µL)	1.14 µl	56.82 µl

## PROCEDURE

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

- 1 | Induce GAL4 activation of the TaDa cassette 16-24 hours prior to collection  
**CRITICAL STEP:** A Dam-only control, driven under the same experimental conditions as the Dam-fusion protein, is always required to be processed in parallel.
- 2 | Dissect tissue in PBS. Collect enough material to include at least 15,000 labelled cells; include more labelled cells if expression of the fusion-protein was induced for less than 16 hours.
- 3 | When enough cells have been collected, remove excess PBS and store until required.

PAUSE POINT: samples can be stored at -20°C or -80°C until required

### Extraction of genomic DNA (using Qiagen DNA Micro Kit) • TIMING 1 h 30 min to 2 h

- 4 | This step can be performed using Option A or Option 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 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 freezer and add 175  $\mu$ l 1X PBS to eppendorf tube containing sample. If processing samples containing gut tissue, we recommend the addition of EDTA in order to prevent fragmentation of genomic DNA by nucleases.
- ii. Add 20  $\mu$ l RNase (12.5 $\mu$ g/ml) and pipette mix.
- iii. Use sterile pestle (washed in 100% ethanol) to homogenise sample in PBS.
- iv. Add 20  $\mu$ l Proteinase K (Qiagen DNA Micro Kit) and mix gently; leave for 1 min at room temperature (RT; 22°C).
- v. Add 200  $\mu$ l Buffer AL (Qiagen DNA Micro Kit) and mix gently; incubate at 56°C for 10 mins.
- vi. Cool to room temperature, add 200  $\mu$ l 100% ethanol and mix gently.

**B) Small volumes (<10 $\mu$ l total volume) of dissected tissue TIMING 1 h – 6 h**

- i. Take cut larvae or dissected tissue from freezer. If processing samples containing gut tissue, we recommend the addition of EDTA in order to prevent fragmentation of genomic DNA by nucleases.
  - ii. Add 180 $\mu$ l ATL buffer (Qiagen DNA Micro Kit)
  - iii. Add 20 $\mu$ l Proteinase K (Qiagen DNA Micro Kit) and mix gently.
  - iv. Incubate at 56°C on heat block until digested; flick gently to mix occasionally.  
 ▲ **CRITICAL STEP** Digestion will typically take one hour for most tissues, but may require longer if larger volumes of material are used.
  - v. Cool to room temperature and add 20  $\mu$ l RNase (12.5 $\mu$ g/ml)
  - vi. Incubate at room temperature for 2 mins
  - vii. Mix 200 $\mu$ l Buffer AL (Qiagen DNA Micro Kit) and 200 $\mu$ l 100% EtOH together (per sample) in a separate tube; mix well by vortexing
  - viii. Add 400 $\mu$ l of Buffer AL/EtOH mix to each sample; mix well by inversion.
- 5 | Add all solution (from step 4A(vii) or 4B(viii)) to a spin column (Qiagen DNA Micro Kit).
  - 6 | Spin at >6000 x g for 1 min, discard flow-through and collecting tube.
  - 7 | Add 500  $\mu$ l AW1 solution (Qiagen DNA Micro kit) and spin at >6000 x g for 1 min, discard flow-through and collecting tube.
  - 8 | Add 500  $\mu$ l AW2 (Qiagen DNA Micro kit) solution and spin at >6000 x g for 1 min, discard flow-through and collecting tube.
  - 9 | Spin at 20,000 x g for 3 mins to dry the column.
  - 10 | Transfer spin column to a new 1.5 ml tube, add 50  $\mu$ l AE buffer (Qiagen DNA Micro kit) and leave at RT for at least 30 min. Spin at >6000 x g for 1 min and keep the flow-through.
  - 11 | Run 1  $\mu$ l on a 0.8% agarose gel to check quality. The genomic DNA should be visible as a single band at the top of the gel and not a smear. (TROUBLESHOOTING)

**Digestion of methylated DNA • TIMING ~ 16 hrs**

- 12 | Transfer 43.5 $\mu$ l to a new 1.5ml tube.
- 13 | Add 5 $\mu$ l NEB CutSmart Buffer and 1.5  $\mu$ l DpnI enzyme, mix very gently with either flicking or a P1000 and digest overnight (at least 12 hours) at 37°C.  
 ▲ **CRITICAL STEP** Do not vortex the DpnI solution as this can result in shearing of the genomic DNA.

14 | Clean-up DpnI digested DNA with Qiagen PCR Purification kit according to the manufacturer's instructions; elute in 32  $\mu$ l H<sub>2</sub>O.

■ **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 hrs**

15 | Measure DNA concentration on Qubit or Nanodrop and use either 15  $\mu$ l of undiluted sample (if less than a total of 750 ng) or dilute with H<sub>2</sub>O to a total of 750 ng in 15  $\mu$ l 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  $\mu$ l of sample to 0.2 ml PCR tubes and add 4  $\mu$ l of pre-made Adaptor Ligation Buffer (see REAGENT SETUP).

17 | Add 1  $\mu$ l (400U) T4 DNA ligase enzyme and mix well.

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

#### DpnII digestion of DNA • **TIMING 2 hrs**

19 | Add 19  $\mu$ l of pre-made TaDa DpnII Digestion Buffer (see REAGENT SETUP).

20 | Add 1  $\mu$ l DpnII enzyme and mix well.

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

#### PCR amplification • **TIMING 2.5 hrs**

22 | Add 118  $\mu$ l of pre-made DamID PCR Buffer (see REAGENT SETUP)

23 | Add 2  $\mu$ l Advantage cDNA polymerase enzyme and mix well.

24 | Split the reaction into 4 x 40  $\mu$ l reactions in 0.2 ml PCR tubes.

▲ **CRITICAL STEP** For an efficient PCR amplification, the PCR reaction volume should be kept low ( $\leq$  40  $\mu$ l). 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 sec	65°C 5 min	68°C 15 min
2-4	94°C 30 sec	65°C 1 min	68°C 10 min
5-21	94°C 30 sec	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 PCR adaptors • **TIMING ~16 hrs**

- 26 | Run 3  $\mu\text{l}$  of the PCR product on a 0.8% agarose gel to check quality. A successful amplification should show a smear concentrated between 400bp and 2kb; some unique bands may also be present depending on the particular binding characteristics of the Dam-fusion protein. See Figure 3 for an example.
- 27 | Use Qiagen PCR purification kit according to the manufacturer's instructions to purify the rest of PCR product, use 32  $\mu\text{l}$  H<sub>2</sub>O to elute (leave water on column for at least 5 min before final spin).
- 28 | Measure concentration of purified DNA using a Qubit or Nanodrop and dilute 2  $\mu\text{g}$  in 90  $\mu\text{l}$  H<sub>2</sub>O in a 1.5ml tube. (TROUBLESHOOTING)
- 29 | Add 10 $\mu\text{l}$  NEB CutSmart Buffer and mix well.
- 30 | Sonicate DNA. Sonication time and conditions are highly dependent upon the equipment used, and should be optimised to generate an average DNA fragment size of ~300 bp. (With a Bioruptor Plus, we use 5 min, high power, 30 sec on / 30 sec off.)
- 31 | Check sonicated DNA on an Agilent TapeStation or Bioanalyzer. The fragments should have a mean size of ~300bp (see Fig. 3 for an example).
- 32 | Remove DamID adaptors from sonicated DNA by adding 1 $\mu\text{l}$  of AlwI (mix well). Incubate overnight at 37°C. (TROUBLESHOOTING)
- 33 | Transfer 70 $\mu\text{l}$  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 clean-up • TIMING 40 min**

- 34 | Add 105  $\mu\text{l}$  of AMPure XP beads to 70  $\mu\text{l}$  sample and mix well.  
▲ **CRITICAL STEP** Ensure that the bead stock is fully mixed before pipetting (use vortex)
- 35 | Incubate at RT for 10 min.
- 36 | Place on magnetic stand for 10 min (or until clear) and keep tubes on magnetic stand for the following steps 37-39.
- 37 | Remove and discard supernatant  
▲ **CRITICAL STEP** Do not touch the bead pellet with the pipette tip
- 38 | Wash the beads twice in 190  $\mu\text{l}$  80% EtOH (leave beads on magnet; 30 secs wash time).
- 39 | Remove as much liquid as possible and leave for 5 min to air dry.
- 40 | Remove PCR strip from magnetic stand and resuspend beads in 25  $\mu\text{l}$  Resuspension Buffer.
- 41 | Incubate for 2 min at RT.
- 42 | Place on magnetic stand for 5 min (or until clear).
- 43 | Put 22.5  $\mu\text{l}$  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 $\mu\text{l}$  to measure sample concentration on a Qubit or similar fluorimeter.
- 45 | Dilute sample to (no more than) 500ng of DNA in 20 $\mu\text{l}$  Resuspension Buffer (remainder can be discarded).

**Sequencing library preparation: End repair • TIMING 55 min**

- 46 | Add 7.5 $\mu\text{l}$  End Repair Buffer
- 47 | Add 2.5  $\mu\text{l}$  End Repair Enzyme Mix and mix well.

- 48 | Incubate for 30 min at 30°C.  
49 | Heat inactivate enzymes for 20 mins at 75°C  
■ **PAUSE POINT** End repaired DNA can be stored for up to 6 months at -20°C.

**Sequencing library preparation: Adenylate 3' Ends • TIMING 35 min**

- 50 | Add 0.75 µl Klenow 3' – 5' exo- enzyme.  
51 | Incubate at 37°C for 30 min.

**Sequencing library preparation: Sequencing adaptor ligation • TIMING 35 min**

- 52 | Add 2.5µl of relevant sequencing adaptor.  
▲ **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 filter. In these case, the preferred indices to use (in order) are: 4, 7, 6 and 8.  
53 | Add 2.5 µl NEB Quick ligase enzyme  
54 | Incubate at 30°C for 10 min.  
55 | Add 5µl 0.5M 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 clean-up • TIMING 60 min**

▲ **CRITICAL** Two rounds of bead clean-up (steps 56-65 and 66-75) are required to ensure the complete removal of adaptor dimers

- 56 | Add 41 µl AMPure beads and mix well.  
57 | Incubate at RT for 10 min.  
58 | Place on magnetic stand for 5 min (or until clear).  
59 | Remove 80 µl of supernatant.  
60 | Wash twice in 200µl 80% EtOH (leave beads on magnet; 30 secs wash time).  
61 | Leave for 5 min to air dry.  
62 | Remove PCR strip from magnetic stand and resuspend the beads in 52.5 µl Resuspension Buffer.  
63 | Incubate for 2 min at RT.  
64 | Place on 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 clean-up (steps 66-75).  
■ **PAUSE POINT** Purified DNA can be stored for up to 6 months at -20°C.  
66 | Add 50 µl AMPure beads.  
67 | Incubate at RT for 10 min.  
68 | Place on magnetic stand for 5 min (or until clear).  
69 | Remove 95 µl of supernatant.  
70 | Wash twice in 200µl 80% EtOH (leave beads on magnet; 30 secs wash time).  
71 | Leave for 5 min to air dry.  
72 | Remove PCR strip from magnetic stand and resuspend the beads in 22.5 µl Resuspension Buffer.  
73 | Incubate for 2 min at RT.  
74 | Place on 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: Enrich DNA fragments • TIMING 25 - 40 min**

76 | Add 5 µl PCR Primer Mix.

77 | Add 25 µl PCR Master Mix.

78 | Perform PCR cycle:

Cycle number	Denature	Anneal	Extend
0	98°C 30 sec		
1-6	98°C 10 sec	60°C 30 sec	72°C 30 sec
7			72°C 5 min

▲ **CRITICAL STEP** Greater than 6 cycles may need to be performed if 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 (see Troubleshooting).

**Sequencing library preparation: DNA clean-up • TIMING 30 min**

79 | Add 50 µl AMPure beads to the 50 µl PCR sample.

80 | Incubate at RT for 10 min.

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

82 | Remove 95 µl.

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

84 | Leave for 5 min to air dry.

85 | Remove PCR strip from magnetic stand and resuspend the beads in 32.5 µl Resuspension Buffer.

86 | Incubate for 2 min at RT.

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

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

**Sequencing library quality control • TIMING 60 min**

89 | Check DNA on Agilent Tapestation or Bioanalyzer (see Anticipated Results) and note average fragment size.

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

91 | Pool libraries to final recommended concentration for Illumina sequencing. (DNA molarity =  $\frac{1500}{(\text{fragment size in bp})} \times (\text{concentration in ng/}\mu\text{l})$ ). Ensure that all libraries have the same final concentration when pooled.

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 significantly more reads per library, depending upon genome size.

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

### 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 normalisation 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 localised to the nucleus, and this background methylation is used as the basis of normalisation.

We have developed `damidseq_pipeline`<sup>7</sup>, an automated pipeline for processing samples that is freely available online at ([http://owenjm.github.io/damidseq\\_pipeline/](http://owenjm.github.io/damidseq_pipeline/)) along with detailed usage instructions. The pipeline handles automatic alignment, extension, normalisation and background reduction of the generated sequencing reads; files for human, mouse and fly genomes are provided on the website, and utilities 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 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](http://github.com/owenjm/find_peaks)). In the specific case of RNA pol II datasets, 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  $FDR < 0.01$ .

An example of the final output, illustrating the DNA binding profile of RNA pol II in *Drosophila* larval neural stem cells<sup>7</sup>, is shown in Fig. 4.

End of Box 1

### Timing

A schematic of the Procedure and timeline can be found in Figure 2.

Day 1: Steps 1-13.

Day 2: Steps 14-25.

Day 3: Steps 26-32.

Day 4: Steps 33-75

Day 5: Steps 76-91

### Troubleshooting

Troubleshooting advice can be found in Table 3.

**Table 3** | Troubleshooting table

Step	Problem	Possible reason	Solution
11	Insufficient genomic DNA	Insufficient starting material	Use more starting material (step 2) Although not optimal, as little as 1



		Sample not homogenised enough	$\mu$ g of genomic DNA can be used. Perform more vigorous homogenisation (step 4)
28	Insufficient PCR product	Small number of cells expressing the Dam-fusion protein, relative to the rest of the tissue	Increase starting material at step 2 or the length of GAL4 induction time at step 1.
32	DNA fragment sizes too large	Insufficient sonication	Re-optimize sonication time (step 30)
87	Secondary peak in sequencing library	Exhaustion of PCR in 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 6 cycles.
87	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 genome	Contamination Failure to remove initial adaptors with AlwI  Adaptor dimers or adaptor concatemers present in sequencing library	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 filter.  Adaptor dimers or concatemers will be visible in 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 pipette accuracy.

## Anticipated results

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

The PCR amplification should produce DNA fragments ranging from approximately 400 bp to 2500 bp (see Figure 3A,B). The total yield may range from 2 to 20  $\mu$ g, depending on the number of labelled 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 ( $\pm$  50 bp). See Figure 3A,C for an example. DamID adaptors are subsequently removed with AlwI digestion.

#### **Sequencing library preparation (Steps 34-90)**

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

We assess overall quality through the percent 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<sub>2</sub> ratio profile from a DamID-seq experiment should show strong signal for bound regions with low background (see Figure 4 for an example profile using RNA pol II.) If profiling a transcription factor, peaks should be detectable through software (see Box 1); if profiling transcription via RNA pol II, genes with an FDR < 0.01 should be detectable following software analysis (see Box 1).

## Figure Legends

**Figure 1** | TaDa can profile the interaction of any protein with DNA or chromatin in a cell type-specific manner. For example, RNA polymerase II (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 enzyme such as Polycomb (Pc).

**Figure 2** | Flow chart for processing of TaDa material, with protocol steps listed. The entire procedure is completed in under five days.

**Figure 3** | Sample TapeStation plots showing the expected fragment size distribution following TaDa. (A) Gel image of samples following the TaDa PCR step (step 29) and sonication (step 35). (B) Electropherogram of the TaDa PCR sample; (C) electropherogram of the sample post-sonication. The sample illustrated was prepared by driving the DAM-only construct in adult neurons for 24 hours, with DNA isolated from whole adult heads.

**Figure 4** | Example log<sub>2</sub> ratio profile of RNA pol II binding in *Drosophila* neural stem cells. DNA was isolated from 50 *Drosophila melanogaster* larvae for both Dam-polII and Dam-only samples; larvae were torn in half and the anterior halves (which contain the brain) 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<sup>ts</sup> and induced for 16 hrs at 29°C from 72 hrs 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.

## Author contributions

OJM, TDS and AHB designed the research and analysed the data; TDS and OJM developed the protocols; OJM, TDS and AHB wrote the paper.

## Acknowledgements

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).

## Competing financial interests

The authors declare that they have no competing financial interests.

## References

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**, 101–12 (2013).
2. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–8 (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**, 401–15 (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**, 3371–3 (2015).
8. Fillion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–24 (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., Letarnek, 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**, 948–51 (2008).
13. Vogel, M. J. *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.* **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**, 978–94 (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**, 271–82 (2013).
21. Otsuki, L., Cheatham, S. W. & Brand, A. H. Freedom of expression: Cell-type-specific gene profiling. *Wiley Interdisciplinary Reviews: Developmental Biology* **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**, 4818–25 (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. U. S. A.* **98**, 12596–12601 (2001).
32. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–46 (2012).