The FlyORF library of fly lines (Bischof *et al.* 2013) contains inducible ORFs that cover 74% (560/757) of all known and predicted TFs and chromatin-associated proteins in the *Drosophila* genome (Bischof *et al.* 2013, 2014). The library was designed with cassette exchange features, presenting the possibility of converting existing FlyORF lines into TaDa lines to profile DNA binding via a simple cross. Here, we describe FlyORF-TaDa, a new system that allows the conversion of FlyORF lines to TaDa lines via Flippase-mediated cassette exchange. FlyORF-TaDa permits the rapid and easy generation of new TF-binding profiles without cloning or the creation of transgenic animals, with recombinants easily identified by eye color and fluorescence.

Materials and methods

Fly husbandry

Drosophila were raised on media containing 5% (w/v) yeast, 5.5% (w/v) dextrose, 3.5% (w/v) cornflour, 0.75% (w/v) agar, 0.25% (v/v) Nipagin, and 0.4% (v/v) propionic acid. Flies were grown in incubators at 70% humidity on a 12-h/12-h light/dark cycle.

Expression constructs

The *pFlyORF-TaDa* vector was created by cutting *pTaDaG2* (Delandre *et al.* 2020) with BglII/NotI (NEB) and inserting a synthetic gBlock (IDT) containing an *FRT5* site followed by an in-frame stop codon via NEB HiFi Assembly (NEB) to create *pTaDaG2-FRT5*. *pTaDaG2-FRT5* was cut with AfeI/BstBI and a synthetic gBlock (IDT) containing *3xP3-DsRed2-small_t_intron-polyA* via NEB Hifi Assembly to generate *pTaDaG2-FRT5-3xP3-dsRed2*. Finally, mini-white (including the 240 bp upstream and 630 bp downstream *white* regulatory regions) was amplified via PCR from *pTaDaG2* and inserted into *pTaDaG2-FRT5-3xP3-dsRed* cut with XhoI and XbaI via NEB Hifi Assembly to generate *pFlyORF-TaDa*.

pTaDaG2-MRG15 was generated by inserting a synthetic gBlock (IDT) containing the sequence of MRG15-RA into the *pTaDaG2* vector cut with XbaI/XhoI. All plasmids were sequence-verified via Sanger sequencing (ABI). Plasmid maps were generated using SnapGene software (Insightful Science).

Fly lines

The *worniu-GAL4;tub-GAL80ts* line (Albertson *et al.* 2004) was used as a driver for neural stem cells. The *FlyORF-TaDa* line was created through phiC31-integrase-mediated insertion into *ZH-86FB* (the site used by the FlyORF expression library) on chromosome 3, by injecting *pFlyORF-TaDa* at 200 ng/µl into embryos from a *y[1],M{RFP[3xP3.PB] GFP[E.3xP3]=vas-int.Dm}ZH-2A,w[*];M{3xP3-RFP.attP}ZH-86Fb* (Bloomington # 24749) in which the landing site *3xP3-RFP* marker had been previously removed via Cre-mediated recombination. The resulting transgenic line was backcrossed three times to w^{1118} before a homozygous *w;+;FlyORF-TaDa* stock was generated. The *w;hs-FlPD5;FlyORF-TaDa* was generated by crossing to w^{1118};*P{y[+t7.7] w[+mC]=hs-FLPD5}attP40/CyO* stock (Bloomington # 55814). The *TaDaG2-MRG15* line was generated by BestGene, Inc (CA), through phiC31-integrase-mediated insertion of *pTaDaG2-MRG15* into *attP2* on chromosome 3L. *TaDaG2-Dam* flies were used as previously published (Delandre *et al.* 2020).

FlyORF cassette exchange

Homozygous *hs-FlpD5;FlyORF-TaDa* virgin females were crossed to males from *FlyORF* lines. Progeny (third instar, 96 h after larval hatching) were heat shocked at 37°C for 60 min. After eclosion, F1 male flies were crossed to TM3/TM6B virgin females. F2 males and virgin females with TM6B and exhibiting the correct eye phenotype (*w-;3xP3-dsRed2+*) were crossed to establish a balanced stock.

Targeted DamID

Appropriate lines (for conventional TaDa: *TaDaG2-MRG15* and *TaDaG2-Dam*; for FlyORF-TaDa: *FlyORF-TaDa-MRG15* and *FlyORF-TaDa* lines) were crossed to *worniu-GAL4;tub-GAL80ts* virgin females in cages. Embryos were collected on apple juice agar plates with yeast over a 4-h collection window and grown at 18°C for 2 days. Newly hatched larvae were transferred to food plates for a further 5 days at 18°C, before shifting to 29°C for 24 h.

Larval brains were dissected in PBS, and processed for DamID-seq as previously described (Marshall *et al.* 2016; Marshall and Brand 2017) with the following modifications. Briefly, DNA was extracted using a Quick-DNA Miniprep plus kit (Zymo), digested with DpnI (NEB) overnight, and cleaned up with a PCR purification kit (Machery-Nagel), DamID adaptors were ligated, digested with DpnII (NEB) for 2 h, and amplified via PCR using MyTaq DNA polymerase (Bioline).

Following amplification, 2 µg DNA was sonicated in a Bioruptor Plus (Diagenode). DamID adaptors were removed by AlwI digestion, and 500 ng of the resulting fragments end-repaired with a mix of enzymes [T4 DNA ligase (NEB) + Klenow Fragment (NEB) + T4 polynucleotide kinase (NEB)], A-tailed with Klenow 3′→5′ exo- (NEB), ligated to Illumina Truseq LT adaptors using Quick Ligase enzyme (NEB), and amplified via PCR with NEBNext Hi-fidelity enzyme (NEB). The resulting next-generation sequencing libraries were sequenced on a HiSeq 2500 (Illumina).

Bioinformatic analyses

DamID-binding profiles were generated from NGS reads using damidseq_pipeline (Marshall and Brand 2015) and visualized using pyGenomeTracks (Ramírez *et al.* 2018). Peaks were called using a three-state hidden markov model via the hmm.peak.caller R script (freely available from https://github.com/owenjm/hmm.peak.caller). Heatmaps were generated via the ComplexHeatmap R package (Gu *et al.* 2016).

Gene ontology (GO) enrichment plots were performed using the ClusterProfiler R package with Bonferroni–Holm-adjusted *P*-values; enrichmap plots were generated by limiting GO terms to <1000 genes and using the simply() function to remove redundancy (Yu *et al.* 2012).

MRG15 enrichment by chromatin state (significance and odds ratio) was assessed via Fisher's exact test with an alternate hypothesis of "greater", for contingency tables of genomic coverage of MRG15 *vs* genomic coverage of the chromatin state in neural stem cells (NSCs). All *P*-values were Bonferroni–Holm adjusted. All other plots were generated using R (R Core Team, 2020).

Data availability

The sequence of *pFlyORF-TaDa* is available under Genbank accession number MT733231, and the plasmid DNA is available upon request. DamID-seq data and processed bedgraph and analysis files have been deposited in NCBI GEO, accession number GSE159632. The *w;+;TaDaG2-MRG15*, *w;+;TaDaG2-Dam*, and *w;+;FlyORF-TaDa-MRG15* fly lines are available upon request; the *w;hsFlp-D5;FlyORF-TaDa* and *w;+;FlyORF-TaDa* lines have been deposited with the Bloomington Drosophila Stock Center.

Supplementary material is available at figshare DOI: https://doi.org/10.25387/g3.13245053.

Results and discussion

Design of the FlyORF-TaDa system

The FlyORF library of inducible ORFs incorporates an FRT5 site immediately upstream of each ORF, allowing Flippase-mediated exchange with an upstream donor cassette in the same genomic insertion site (Bischof *et al.* 2013). To convert these lines to TaDa lines, FlyORF-TaDa provides the UAS-inducible bicistronic transcript of TaDa vectors upstream of an FRT5 site as the donor cassette. FlyORF-TaDa uses the TaDaG2 version of the TaDa bicistronic transcript, in which the primary ORF is myristoylated GFP, providing the advantage of easily-detectable fluorescent labeling of cells being profiled within experimental samples (Delandre *et al.* 2020).

The bicistronic cassette is followed by an *FRT5* site in frame with *Dam* (Figure 1A and Supplementary Figure S1A). Without recombination, a stop codon positioned directly after the FRT5 site allows the FlyORF-TaDa line to be used as a Dam-only control for DamID signal normalization (Supplementary Figure S1B). Following recombination with a FlyORF line, the FRT5 site together with the Gateway cloning residual attB1 site present as part of the FlyORF library construction (Bischof *et al.* 2013) becomes a 23-amino-acid protein linker region (Supplementary Figure S1C).

To prevent the need to PCR-screen progeny, the vector was additionally designed with a *3xP3-dsRed2* eye marker upstream of the TaDa cassette, and the *mini-white* eye marker downstream of the *FRT5* site (Figure 1A and Supplementary Figure S1A). While both parental strains used for a conversion are *w+*, upon Flippase-mediated strand exchange, successful recombinants will be *w-,dsRed2+* making screening of recombinants by eye color fast and straightforward. [We note that the FlyORF lines contain a *3xP3-RFP* marker as part of the landing site (removed in the FlyORF-TaDa lines) making screening for recombinants on eye fluorescence alone difficult, although we observe significantly more intense eye fluorescence from the *3xP3-dsRed2* marker.]

The *pFlyORF-TaDa* vector was inserted into ZH-86FB, the same insertion site used by the FlyORF library, and the resulting line was crossed with a heat-shock-inducible Flippase (*hs-FlpD5*) line to yield an *hs-FlpD5;FlyORF-TaDa* donor line. TaDa alleles are generated by crossing any FlyORF line to this donor line, in conjunction with a heat-shock of progeny during the larval phase. Crossing the resulting adult males to virgin females with chromosome 3 balancers yields a typical recombination frequency of 50% (combined totals for two separate FlyORF-TaDa conversions: 7 vials with at least one recombinant/14 total vials scored with a minimum of 50 progeny). An illustration of the crossing scheme is shown in Supplementary Figure S2.

Profiling of MRG15 binding in neural stem cells using the FlyORF-TaDa system

To determine whether FlyORF-TaDa lines generated through the system faithfully profiled binding in a cell-type-specific manner, we obtained binding profiles for the chromatin-binding protein MRG15 in NSCs. We obtained profiles from a *FlyORF-TaDa-MRG15* and compared these to MRG15 profiles generated using conventional TaDa, driving expression in both cases with the NSC-specific driver, *worniu-GAL4*. Excellent correlation between samples was observed throughout (Figure 1B), both between individual replicates (Supplementary Figure S3) and when comparing the average profiles from the two systems (Figure 1B). In particular, the two biological replicates generated using the FlyORF-TaDa system showed extremely high reproducibility (Pearson's correlation between replicates, 0.95; Supplementary Figure S3).

MRG15 is a protein associated with the H3K36me3 histone mark (Zhang *et al.* 2006), which in turn is associated with transcribed exons (Kolasinska-Zwierz *et al.* 2009; Kharchenko *et al.* 2011). The protein is also a key component of the "Yellow" active chromatin state described in flies by Filion *et al.* (2010), a study that reduced chromatin configurations in *Drosophila* to five broad classification types, or colors. In concordance with these observations, we found MRG15 bound at previously published Yellow chromatin domains in NSCs (Marshall and Brand 2017) (Figure 2A), with MRG15-bound peaks covering 80% (9.8 Mb of 12.2 Mb) of Yellow chromatin. While MRG15 occupancy was significantly enriched to some degree over all three active chromatin states in NSCs, by far the greatest enrichment was observed over Yellow chromatin [Figure 2B; $\log_e\theta$ (Yellow chromatin) = 2.91, Fisher's exact test].

Genes covered by bound peaks were enriched for GO functions associated with metabolism (Figure 3 and Supplementary Figure S4), a known association for genes covered by Yellow chromatin, both in cell lines (Filion *et al.* 2010; van Bemmel *et al.* 2013) and in NSCs (Marshall and Brand 2017). Importantly, we also observed an enrichment for genes involved in nervous system development expressed in NSCs (Figure 3), consistent with occupancy over transcribed exons and indicating that the FlyORF-TaDa system can faithfully profile cell-type-specific protein binding *in vivo*.

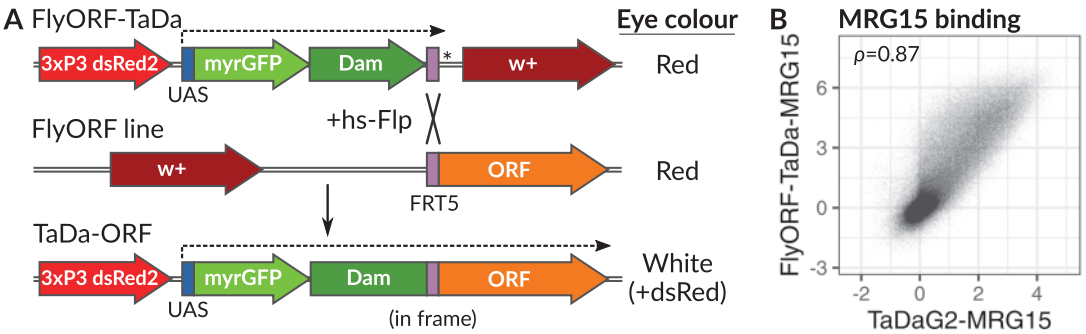

Figure 1. The FlyORF-TaDa system generates new lines for cell-type specific profiling. (A) Schematic representation of the FlyORF-TaDa system. When the FlyORF-TaDa donor line is crossed to a FlyORF line and a heat-shock-inducible *flippase* (hs-Flp), recombination between compatible FRT5 sites generates a new TaDa line for the ORF in the progeny. These lines can be screened for in the subsequent generation by eye color. Dashed lines show GAL4/UAS-induced transcription; *stop codon. (B) Correlation of DNA-binding profiles generated for MRG15 in NSCs using the FlyORF-TaDa system, compared to conventional TaDa (Pearson's correlation is shown).

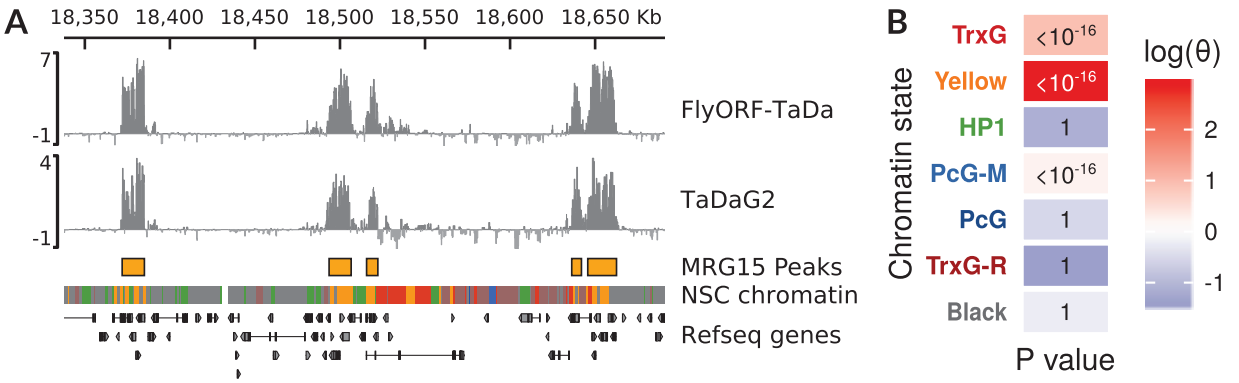

Figure 2. Characteristics of MRG15 binding in NSCs. (A) MRG15 binds to regions of Yellow chromatin in NSCs. Profiles obtained via FlyORF-TaDa and conventional TaDa are shown; peaks were called on the average binding profile of the combined MRG15 datasets. NSC chromatin state data and colors are from Marshall and Brand (2017). A 300-kb region of chrX is illustrated. (B) Significance and odds ratio (θ) of enrichment of MRG15 peak occupancy within NSC chromatin states as assessed via Fisher's exact test.

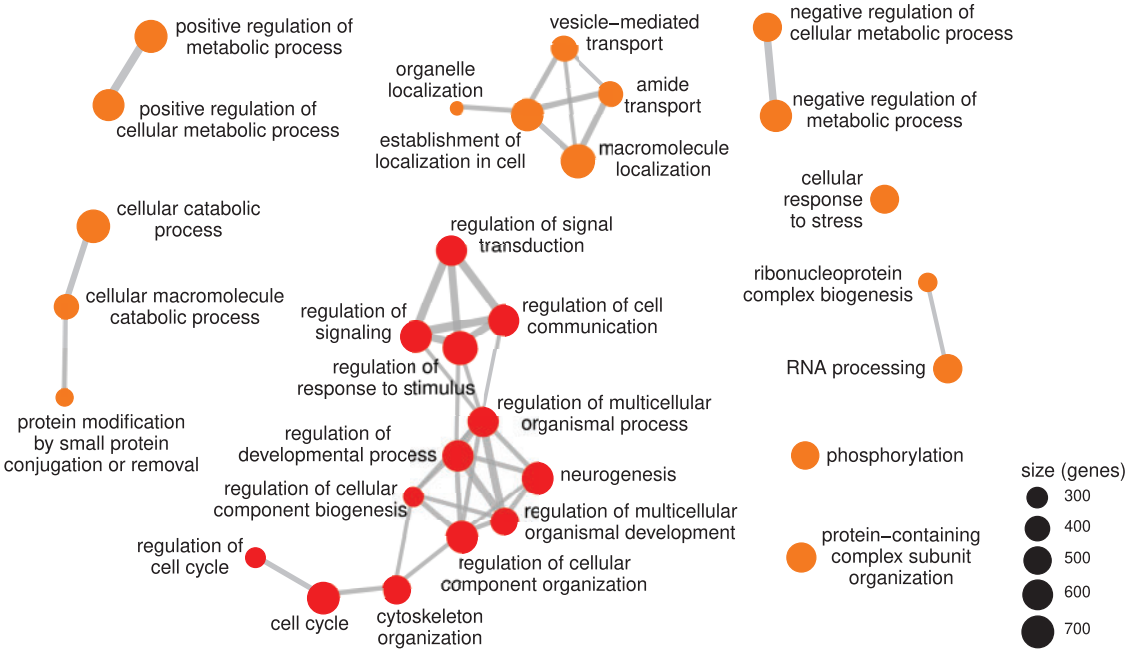

Figure 3. GO term enrichment for genes covering MRG15-bound peaks identifies terms for both metabolism and developmental neurogenesis (non-redundant GO terms covering <1000 genes illustrated). Metabolism-related terms are colored in orange; developmental terms are colored in red. All terms illustrated were significantly enriched (all adjusted *P*-values $<10^{-28}$).

Limitations

Although versatile, some potential limitations of the FlyORF-TaDa system should be considered when using the converted lines for profiling. In particular, the FlyORF lines that are convertible to TaDa lines (*i.e.* lines that include an FRT5 site) all incorporate a long (22aa) flexible linker combined with 3xHA tags at the C-terminal. There is some evidence that the presence of HA tags may disrupt some aspects of protein function in unusual cases. For the FlyORF library, the presence of these tags was shown—albeit in overexpression experiments—to cause a gain of function in 11% of lines, and a failure to cause a phenotype in 22% of lines, when compared to overexpression of the untagged variant (Bischof *et al.* 2013). A related study of GFP- and FLAG-tagged proteins from the FlyFos library of lines suggested that 33% of tagged lines were unable to rescue the corresponding deficiency, although this failure was ascribed to the lack of long-range regulatory elements of complex developmental TFs rather than the presence of tags themselves (Sarov *et al.* 2016). It is unknown whether in any of these cases protein–DNA binding was affected. Nevertheless, it is possible that in a minority of cases the presence of the HA tags may lead to unrepresentative binding profiles.

Another consideration is that the FlyORF-TaDa system only generates fusions with Dam fused to the N-terminus of the ORF protein (again via a long linker generated via the FRT5 site). It remains unclear as to whether Dam can disrupt TF binding in some circumstances, although previous studies have shown that Dam-fusion proteins yield binding profiles broadly comparable to ChIP-seq data obtained for the same protein (Cheetham *et al.* 2018; Tosti *et al.* 2018; Szczesnik *et al.* 2019).

Notably, the FlyORF libraries also include an FRT2 site immediately 3′ of the ORF. Both the removal of the C-terminal linker and 3xHA tags, and the generation of C-terminal Dam-fusion proteins, would in the future be possible with different donor lines designed on a similar basis to the system presented here.

Conclusion

The FlyORF-TaDa system places fast and straightforward cell-type-specific profiling of TF binding within the reach of any fly lab, allowing the profiling of over 74% of all TFs and chromatin-associated proteins via a simple genetic cross. Establishment of new recombinant lines from the parental FlyORF stocks is achieved in two generations, without the need for cloning, sequencing or transgenesis. The donor *w;hs-FlpD5;FlyORF-TaDa* and Dam-only control *w;+;FlyORF-TaDa* lines have been deposited with, and will be available from, the Bloomington Drosophila Stock Center (BDSC).

With three quarters of all TFs covered by donor FlyORF lines, comprehensive wide-scale cell-type-specific profiling of TF-binding networks can be achieved. TaDa is ideally suited to profiling such networks without introducing the variability of different fixation and antibody isolation steps found in alternative methods such as ChIP-seq. The FlyORF-TaDa system now eliminates a major obstacle to this approach by dramatically reducing the time and costs required to generate the lines required to profile TF networks. To further this aim, we have successfully generated 17 new FlyORF-TaDa lines using the system and are creating a library of converted TF and chromatin factor lines that will be deposited in stages at the BDSC. We anticipate that this resource will prove highly useful to the *Drosophila* transcription and chromatin, and developmental biology communities.

Acknowledgment

We thank G. Jefferies for technical support.

Funding

This work was supported by National Health and Medical Research Council (NHMRC) Australia grants APP1128784 and APP1185220 to O.J.M., a Wellcome Trust Investigator grant 104567 to T.D.S., and a Biotechnology and Biological Sciences Research Council (BBSRC) UK grant BB/P017924/1 to T.D.S. and G.N.A.

Conflicts of interest: None declared.

Literature cited

Albertson R, Chabu C, Sheehan A, Doe CQ. 2004. Scribble protein domain mapping reveals a multistep localization mechanism and domains necessary for establishing cortical polarity. J Cell Sci. 117:6061–6070. doi:10.1242/jcs.01525.

Aughey GN, Cheetham SW, Southall TD. 2019. DamID as a versatile tool for understanding gene regulation. Development. 146:1–8. doi:10.1242/dev.173666.

Bischof J, Bjorklund M, Furger E, Schertel C, Taipale J, *et al.* 2013. A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development. 140:2434–2442. doi:10.1242/dev.088757.

Bischof J, Sheils EM, Björklund M, Basler K. 2014. Generation of a transgenic ORFeome library in Drosophila. Nat Protoc. 9:1607–1620. doi:10.1038/nprot.2014.105.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 118:401–415.

Cheetham SW, Gruhn WH, van den Ameele J, Krautz R, Southall TD, *et al.* 2018. Targeted DamID reveals differential binding of mammalian pluripotency factors. Development. 145:dev170209. doi:10.1242/dev.170209.

Delandre C, McMullen JPD, Marshall OJ. 2020. Membrane-bound GFP-labelled vectors for Targeted DamID allow simultaneous profiling of expression domains and DNA binding. bioRxiv 1–4. doi:10.1101/2020.04.17.045948.

Doupé DP, Marshall OJ, Dayton H, Brand AH, Perrimon N. 2018. Drosophila intestinal stem and progenitor cells are major sources and regulators of homeostatic niche signals. Proc Natl Acad Sci USA. 115:12218–12223. doi:10.1073/pnas.1719169115.

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, *et al.* 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell. 143:212–224. doi:10.1016/j.cell.2010.09.009.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 32:2847–2849. doi:10.1093/bioinformatics/btw313.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, *et al.* 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature. 471:480–485. doi:10.1038/nature09725.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, *et al.* 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet. 41:376–381. doi:10.1038/ng.322.

Marshall OJ, Brand AH. 2015. Damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. Bioinformatics. 31:3371–3373.

Marshall OJ, Brand AH. 2017. Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. Nat Commun. 8:2271.doi:10.1038/s41467-017-02385-4.

Marshall OJ, Southall TD, Cheetham SW, Brand AH. 2016. Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. Nat Protoc. 11:1586–1598. doi:10.1038/nprot.2016.084.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, *et al.* 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun. 9. doi:10.1038/s41467-017-02525-w.

Reiter F, Wienerroither S, Stark A. 2017. Combinatorial function of transcription factors and cofactors. Curr Opin Genet Dev. 43:73–81. doi:10.1016/j.gde.2016.12.007.

Sarov M, Barz C, Jambor H, Hein MY, Schmied C, *et al.* 2016. A genome-wide resource for the analysis of protein localisation in Drosophila. Elife. 5:1–38. doi:10.7554/eLife.12068.

Southall TD, Gold KS, Egger B, Davidson CM, Caygill EE, *et al.* 2013. 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. doi:10.1016/j.devcel.2013.05.020.

Stampfel G, Kazmar T, Frank O, Wienerroither S, Reiter F, *et al.* 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature. 528:147–151. doi:10.1038/nature15545.

Szczesnik T, Ho JWK, Sherwood R. 2019. Dam mutants provide improved sensitivity and spatial resolution for profiling transcription factor binding. Epigenetics Chromatin. 12:36.doi:10.1186/s13072-019-0273-x.

van Bemmel JG, Filion GJ, Rosado A, Talhout W, De Haas M, *et al.* 2013. A network model of the molecular organization of chromatin in Drosophila. Mol Cell. 49:759–771. doi:10.1016/j.molcel.2013.01.040.

Tosti L, Ashmore J, Tan BSN, Carbone B, Mistri TK, *et al.* 2018. Mapping transcription factor occupancy using minimal numbers of cells in vitro and in vivo. Genome Res. 28:592–605. doi:10.1101/gr.227124.117.

Yu G, Wang LG, Han Y, He QY. 2012. ClusterProfiler: an R package for comparing biological themes among gene clusters. Omi J Integr Biol. 16:284–287. doi:10.1089/omi.2011.0118.

Zhang P, Du J, Sun B, Dong X, Xu G, *et al.* 2006. Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Res. 34:6621–6628. doi:10.1093/nar/gkl989.

Communicating editor: M. Boutros