DOI 10.1002/dvg.22925

Tools for DNA adenine methyltransferase identification analysis of nuclear organization during *C. elegans* development

Running title: DamID tools for *C. elegans*

Rahul Sharma[1,2 φ], Dominic Ritler[1 φ] and Peter Meister[1 +]

[φ] these authors contributed equally

[1] Cell Fate and Nuclear Organization, Institute of Cell Biology, University of Bern, Switzerland

[2] Graduate School for Cellular and Biomedical Sciences, University of Bern, Switzerland

[+] correspondence: peter.meister@izb.unibe.ch

Abstract

C. elegans has recently emerged as a valuable model to understand the link between nuclear organization and cell fate, by combining microscopy approaches, genome-wide mapping techniques with advanced genetics. Crucial to these analyses are techniques to determine the genome-wide interaction pattern of proteins with DNA. Chromatin immunoprecipitation has proven valuable but it requires considerable amounts of starting material. This is sometimes difficult to achieve, in particular for specific genotypes (balanced strains, different sexes, severe phenotypes…). As an alternative to ChIP, DNA adenine methyltransferase identification by sequencing (DamID-seq) was recently shown to be able to characterize binding sites in single mammalian cells. Additionally, DamID can be achieved for cell-type specific analysis by expressing Dam fusion proteins under tissue specific promoters in a controlled manner. In this report, we present a user-friendly pipeline to analyse DamID-seq data in *C. elegans*. Based upon this pipeline, we provide a comparative analysis of libraries generated with different starting material and discuss important library

Received: Dec 14, 2015; Revised: Jan 01, 2016; Accepted: Feb 02, 2016

features. Moreover, we introduce an adaptation of an imaging based tool to visualize *in vivo* the cell-specific tridimensional binding pattern of any protein of interest.

Keywords
C. elegans, DNA adenine methyltransferase identification, DamID, protein-DNA interactions

Introduction

Inside the nucleus, eukaryotic genomes are packaged into chromatin and interact with thousands of proteins. These range from transcription factors to chromatin modifiers and structural proteins shaping the genome in three dimensions. Genome-wide identification of the binding sites for these various proteins has become a classical tool to understand their function in regulating gene expression. Two techniques have been used in nematodes, chromatin immunoprecipitation (ChIP) and DNA adenine methyltransferase identification (DamID, Askjaer *et al.*, 2014).

ChIP relies on *in vivo* crosslinking of DNA with proteins, shearing of the crosslinked chromatin and pull-down with an antibody directed against the protein of interest. ChIP requires a considerable amount of starting material (often millions of animals, Askjaer *et al.*, 2014, 2014). This is hard to achieve when the genotype of interest cannot be easily selected (*e.g.* balanced mutations, specific stages or different sexes). In contrast to ChIP, DamID can be carried out with a limited number of animals (Sharma *et al.*, 2014) or even single mammalian cells (Kind *et al.*, 2015). The DamID technique was developed by Bas van Steensel and Steven Henikoff in yeast and later used in Metazoans including nematodes (Guelen *et al.*, 2008; Pickersgill *et al.*, 2006; Towbin *et al.*, 2012; van Steensel and Henikoff, 2000). Dam is an *E. coli* enzyme that methylates adenines in GATC motifs. These methylated motifs can be identified using a combination of restriction enzyme digestion and PCR amplification of methylated fragments. While the basics of the method have not changed since the original publication, two dramatic improvements to the technique were described in the last years. First, the protocol was optimized to bypass the need for DNA isolation and purification, therefore allowing a single tube protocol and

reducing dramatically the amount of starting material (Kind *et al.*, 2015). Second, the identification of the amplified fragments is no longer carried out by hybridization onto microarrays but by direct high throughput sequencing. The latter resulted in enhanced resolution, sensitivity and accuracy in detection of methylated fragments. More importantly, DamID works efficiently for cell-type specific analysis of protein-DNA interactions. Techniques like Targeted-DamID (TaDa) (Southall *et al.*, 2013) have been described in *Drosophila* and can be easily adapted to *C. elegans* by using a tissue specific promoter for controlled expression of Dam-fusion proteins. Alternatively, a Cre/*lox* recombinase approach based on the tissue specific activation of the *Dam* fusion can be used to initiate *Dam* protein expression solely in the cells of interest.

In this short article, we present an outline of the DamID protocol and an R-based pipeline for the analysis of DamID-sequencing data. Based on this pipeline, we provide a comparison between DamID libraries generated using various starting material in *C. elegans*. Moreover, we present an adaptation to nematodes of the 'methyl-adenine tracer' system, which allows the imaging of subnuclear localization of methylated chromatin in individual nuclei of living animals using microscopy techniques.

Results and Discussion

DamID in *C. elegans*: theoretical considerations

DamID is based on the expression at trace levels of a fusion between the protein of interest and the *E. coli* Dam methyltransferase. Dam methylates adenines in GATC motifs in close proximity to the protein of interest (Figure 1 (a)). The technique relies on cutting these methylated motifs with the methylation-specific *Dpn*I enzyme, ligating double stranded adapters to the cut restriction sites and amplifying the restriction fragments using a primer which hybridizes in the adapter sequence (Figure 1 (b-c)). Previously, these amplicons were hybridized to microarrays, while the newly developed approach is based on preparing a classical high-throughput sequencing library from these amplicons (Figure

1 (d)). In contrast to ChIP where all sequences bound to a given protein are immunoprecipitated and sequenced, DamID relies on discrete motifs in two states, methylated or not. This discrete pattern allows stringent quality controls in the sequencing step because every DamID read should start with a GATC motif. The achievable resolution of the technique however depends on the sequence of the genome (GATC motif density) as well as the successive steps (PCR amplification and sequencing). As the number of motifs in the genome is limited, there is moreover no need to sequence very deep: a million reads is sufficient to provide a genome-wide pattern of binding sequences at low (100 kb) resolution. DamID is particularly well adapted when one wants to describe chromosome or megabase scale interactions (e.g. localization close to the nuclear lamina or nuclear pores, Gonzalez-Aguilera *et al.*, 2014; Guelen *et al.*, 2008; Pickersgill *et al.*, 2006; Sharma *et al.*, 2014; Towbin *et al.*, 2012). However, many profiles were achieved with nuclear proteins which bind smaller regions, such as transcription factors, RNA polymerase, chromatin remodellers or insulators (Filion *et al.*, 2010; Southall *et al.*, 2013). Methylation by any Dam fusion protein requires the GATC motif to be accessible to the enzyme. Therefore, methylation of a given motif depends on both the accessibility status of the region and the localization of this region relative to the Dam fusion protein. To correct for this difference in accessibility, a DamID experiment is carried out with the protein of interest and a freely diffusible Dam fused to GFP (Towbin *et al.*, 2012; Vogel *et al.*, 2007). The methylation profile obtained with the free Dam is used to normalize the methylation profile of the protein of interest for accessibility.

The *C. elegans* genome (WS220) is slightly over 100 megabases and contains 269’036 GATC motifs. The median and mean distance between these motifs in the nuclear genome is 206 bp and 370 bp, respectively (Figure S1 (a)). Along the chromosomes, the distance is fairly constant, with only three regions in which the distance between GATC motifs is too large for these regions to be amplified by PCR, on chromosome II, III and chromosome X (3.3 kb, II:15’275’039-15’278’388; 9.9 kb, III:7’439’384-7’449’353; 32.1 kb, X:4’024’591-4’056’759). Genome-wide only 206 GATC motifs (0.076%) are flanked on

both sides by fragments exceeding 2000 bp, a size most likely not amplified using PCR with pooled fragments. These GATCs are therefore not possible to identify with DamID-seq.

DamID-seq relies on sequencing one (or both) outer end(s) of the amplicons starting with the adapter (Figure 1 (d-e)). For each read, checking adapter and GATC presence at the beginning of the read is an important quality control, as breaks occur during the restriction/digestion/amplification procedure and are ligated to the sequencing adapters. In our pipeline, reads which do not contain the adapter are therefore removed from the library as they do not originate from the DamID procedure. The remaining reads are mapped to the genome, which requires them to be located in a unique and mappable region of the genome (Kind *et al.*, 2015). *In silico*, 95.3% of all 87 bp GATC-flanked sequences are uniquely mappable using RBowtie. As a consequence, the median and mean distances between mappable GATCs slightly increase to a median of 221 bp median and a mean of 401.9 bp (Figure S1 (b)). Two additional wider regions on chromosome V (86 kb, V:8813178-8898875; 217 kb V:2'346'496-2'563'194, WS220) are not mappable due to a high number of repeat elements as well as the left telomere of chromosome IV (48 kb, IV:1-47849, WS220).

An R analysis pipeline for quality control and analysis of DamID-seq data

We constructed an initial analysis pipeline (https://github.com/damidseq/RDamIDSeq) using the free statistical language R (RCoreTeam, 2013). The pipeline is based on a single text file describing the nature of the samples (protein of interest or free Dam normalization) used as a base to run the pipeline. First, sequencing reads in fastq format and their corresponding quality strings are imported into R for each sample. Each read is then scanned for the presence and the position of the PCR primer (DamID adapter) sequence, followed by a GATC motif. Reads missing the primer/GATC sequence are discarded while the DamID adapter sequence is cut from the remaining reads. These cut reads are mapped to the *C. elegans* genome using RBowtie. After mapping, a second

quality check compares the position of the reads relative to genomic GATC motifs, removing reads not starting with this motif (sequencing mistakes) and finally assigning each read to a single motif. Read counts per GATC are then summed for a chosen genomic bin size and normalized by total read number. Finally, for each bin, the ratio between the value obtained for the protein of interest and the free Dam control is calculated and replicas averages are computed.

The script outputs a number of useful quality control metrics related to the sequencing library (fastq read quality), the trimming and mapping procedure as well as the Pearson correlation between the samples and genome wide counts per GATC and per chosen bin size. The number of reads passing each step provides valuable information on the quality of the experiment. If many reads do not contain the adapter, this signals a low sequencing quality and/or the presence of many breaks in the original DNA preparation source. Low sequencing quality or contamination with bacterial DNA will be detected by poor mapping to the genome. In the final comparison step between reads and GATC motifs positions, less than 1% of the reads should be removed as GATC is part of the adapter detection step and this reflects sequencing mistakes.

DamID-seq in *C. elegans*: sequencing depth and resolution

A recurring question with high-throughput sequencing methods is sequencing depth: how many reads are needed to achieve reproducible results? Two key factors appear determinant for DamID-seq: the nature of the protein whose profile is analysed and the desired binning of the genome for the interaction maps.

For *GFP::Dam* libraries, we observe that even for sequencing depth higher than 10^6 DamID reads, the number of detected mappable GATC motifs in the library does not increase over 75-80% of the mappable motifs, whether the library was made using entire animals or purified DNA (for a protocol using DNA as starting material, see Gomez-Saldivar *et al.*, 2016, Figure 2 (a)). This is likely because restriction fragments containing these motifs are too small to get amplified and sequenced. Indeed, the comparison of the

sizes of the fragment located next to GATC motifs absent from the libraries shows a strong bias for small and long fragments (Figure S2). This implies that sequencing deeper than one million reads does provide information on the average methylation frequency of these sites, but not on their localization. Jackpot PCR events do not appear to be a major issue as the GATC counts in samples with different PCR cycles after adapter ligation are very well correlated (Figure S3). In contrast, the Dam fusion protein influences the proportion of recovered GATC motifs: a higher fraction of mappable GATCs can be recovered from *dam::lamin* libraries. Several explanations can be put forward. First, the Dam-lamin fusion protein could be more stable than GFP-Dam, leading to overall increased methylation levels. Second, lamina-associated regions are cell-type specific (see below and Meister *et al.*, 2010). As the DamID profile is carried out using entire animals hence mixed cell types, Dam-lamin accessible GATC motifs may vary between cell types and collectively more GATC motifs can be accessed by the Dam. In any case, several libraries should be made for a given Dam fusion to determine the optimal sequencing depth.

Logically, the number of reads needed to achieve reproducible results is inversely proportional to the size of the genome bins analysed (Figure 2 (b)). When the genome is binned into large stretches (>10 kb), the number of reads needed to achieve good reproducibility is low. In contrast, binning the genome into smaller stretches will require deeper sequencing (Figure 2 (b)). For a given binning of the genome, one can use an empirical approach to quantify whether the libraries were sequenced deep enough. Using all reads from a library, sub-libraries are created *in silico* and the DamID profile obtained with the sub-libraries is correlated with the full library DamID profile. A conservative estimate is that if sequencing depth was sufficient, correlation should be high (>0.95) even if half of DamID reads are removed (Figure 2 (c), black). In contrast, correlation decreases fast when reads are insufficient (Figure 3 (c), green).

DamID-seq is a very sensitive method: a single adult animal is sufficient to generate a profile, which is quite similar to the profile obtained with 20 adults (Figure 2 (d), compare

Dam::lamin 1 and 20 adults). The reads frequency for a single Dam fusion is not really informative and needs to be normalized for accessibility (Figure 2 (e)). The obtained profile is then highly similar to DamID profiles previously obtained with DNA as starting material and microarray hybridization (red line, Towbin *et al.*, 2012). The steps carried out by the script presented are an initial analysis of the quality of the DamID data. Downstream treatment of the data depends very much on the biological question and should be tailor-made. A number of pipelines addressing different questions have been described previously (Kind *et al.*, 2015; Marshall and Brand, 2015).

Single cell observation of chromatin dynamics

DamID relies on the identification of methylated DNA by sequencing or microarray. The current technology in *C. elegans* does not allow single-cell profiling. To overcome this limit, an imaging-based 'molecular contact memory' approach has been described by the Stenseel laboratory, to visualize lamina-associated domains (LADs) in single mammalian cells (Kind *et al.*, 2013). This system combines classical DamID with the expression of a GFP-labelled, enzymatically inactive *Dpn*I fragment. This protein recognizes specifically methylated adenines in GATC motifs and is referred to as the m6A-tracer. The m6A-tracer highlights the methylated genomic regions which were in contact with the dam fusion proteins (lamina, free dam) thus providing a visual readout of past and present methylation events (Figure 3 (a)).

C. elegans is particularly adapted to this approach, as most cells in the nematode are post-mitotic. Indeed, during S phase, DNA replication leads to hemimethylated GATC motifs, which are no longer recognized by *Dpn*I and the m6A-tracer. In post-mitotic animals, DNA methylation accumulates progressively over the entire lifetime of the animal, reflecting chromatin movement in vicinity to the Dam fusion. As a proof of concept, we generated strains which express the m6A-tracer from a single copy insertion under transcriptional control of the *his-72* promoter. For the ease of observation, a *glo-1* mutant genetic background with low autofluorescent gut granules was used. When crossed to a strain expressing a fusion between Dam and a protein-of-interest, the chromatinized sequences

located in close proximity to the Dam fusion protein are modified and bound by the m6A-tracer. The m6A-tracer is very well visible in regions of the animal in which nuclei are densely packed such as the head or the tail. Other cell types are also labeled, except the germline (Figure 3 (b)).

During larval development, changes in the distribution of m6A-tracer signal in lamina versus control strains are observed starting around L3 stage. These differences become prominent from L4 stage onwards, as a reflection of the time needed for the enzymatic reaction – adenine methylation - to occur (Figure 3 (c) and (d)). In a strain expressing a *Dam::lamin* fusion, the fluorescent signal of the m6A-tracer shows peripheral enrichment (Figure 3 (d), Figure S4(a)), suggesting low mobility of the methylated chromatin in those nuclei. The m6A-tracer pattern is quite variable between nuclei, hence lamina-proximal chromatin organization is different between cell types, as suggested by single locus positioning studies (Meister *et al.*, 2010). In contrast, in a strain in which *GFP::Dam* freely diffuses inside the nucleus, the fluorescence is distributed in patches in the entire nuclear volume (Figure 3(c), Figure S4(b). This chromatin is internal, more accessible to *Dam* and located in the middle of the autosomes (as deducted from the DamID profiles). We also observe a certain level of heterogeneity in the signal pattern among different cells (Figure S4, bottom panel), the biological relevance of this heterogeneity remains however unclear.

The future combination of single-cell molecular mapping by DamID-seq with the methyl-tracer method allows one to uncover sequences forming nuclear domains *in vivo*, a major question in the field of nuclear organization.

Material and Methods

Worm strains

N2: wild-type Bristol isolate; BN195 *bqSi195[hsp-16.41p:: dam::myc::lmn-1)] II*; BN196 *bqSi196[hsp-16.41p::gfp::myc::dam] II*; PMW197 *ubsSi11[his-72p::m6At::GFP]* I. *bqSi195[hsp-16.41p::dam::myc::lmn-1] glo-1(zu391)* X.; PMW158 *ubsSi11[his-72p::m6At::GFP]* I. *bqSi195[hsp-16.41p:: dam::myc::lmn-1)] unc-119(ed3)* III. PMW174

bqSi196[hsp-16.41p::gfp::myc::dam] ubsSi11[his-72p::m6At::GFP] I. *unc-119(ed3)* III. All worm strains were routinely cultured on NG2% plates seeded with OP50 at 22.5°C. Strains are available on request.

Sample preparation

Worm strains expressing Dam fusions were grown at 25°C on NG plates seeded with GM48 *dam- E. coli*. Hermaphrodites L4 larvae were picked and transferred to plates without bacteria one hour before pooling 10 animals in a single tube containing 1µl of lysis buffer (10mM TrisAc, 10 mM MgAc, 50 mM KA, 0.67% Tween20, 0.67% Igepal) and snap frozen in liquid nitrogen. Additional 2 µl of lysis buffer with 1mg/ml Proteinase K was added and worms were lysed 4 hours at 42°C, followed by heat inactivation at 95°C for 15'. Methylated GATC were digested with 1U *Dpn*I before ligation of the DamID adapter (van Steensel *et al.*, 2001) and amplification of ligated products by PCR using a single primer (5'-NNNNGTGGTCGCGGCCGAGGATC) complementary to the adapter with Advantage polymerase (Clontech) for 26-32 cycles. 2 tubes (20 worms) were pooled together and the resulting PCR product was purified, end repaired (End-It, Epicentre) and 3'-A-overhangs were added (Klenow fragment 3'->5' exo-). Y-shaped Illumina adapters were ligated and the ligation product was PCR amplified using index-containing primers for multiplexing. Library was purified using AMPure XP 1.8 beads and subjected to HT sequencing in a HiSeq 2500 platform (Illumina, iGE3 Geneva). A bench protocol is provided as supplementary file 1.

Data processing

Read quality is tested using package QuasR in R (version 1.8.4/3.2.2, parameters: bin.len=100000, errors=1, mapping=T, qc=T, species="BSgenome.Celegans.UCSC.ce10", restr.seq="GATC", adapter.seq="CGCGGCCGAG"). Reads missing the adapter (CGCGGCCGAG) followed by the *Dpn*I motif (GATC) were removed. Remaining reads (DamID reads) were cut after the adapter and the remaining sequence mapped to the ce10 *C. elegans* genome using RBowtie (version 1.8.0, parameters: -m 1 --best --strata -S) (Langmead *et al.*, 2009). Once mapped, the number of reads per GATC motif is

counted. For binning, the *C. elegans* genome (ce10) was divided in 100 kb regions and the total GATC read numbers was calculated within this 100 kb regions. The total read number per 100 kb was then normalized using the total number of DamID reads. The ratio between *Dam::lamin*/*GFP::Dam* is calculated for each replica and the mean ratio is plotted using a log2 scale. Sublibraries were generated by reducing the number of reads of the original sequencing file by using Fisher–Yates shuffle (http://homes.cs.washington.edu/~dcjones/fastq-tools/fastq-sample, v1.0). Subsamples were generated from 95% to 5% in 5% steps using the original file as template for each reduction step. Three replicas were made using the following seeds 4357, 4351, 4355. Samples were afterwards run using the R pipeline (see above). The results in the reduced set were then compared against the full set using Pearson correlation. The mean correlation of the full set against the reduced set was plotted for each reduced set. Error bars show minimum and maximum Pearson correlation for the 3 sublibraries.

***In silico* theoretical analysis**

Theoretical reads where made using the ce10 genome by expanding all GATC motifs by 25bp, 60bp and 87bp on both strands. The expanded GATC reads were saved in a fasta file. The adapter sequence was added 5' to all sequences using unix command sed (sed '0~2 s/^/CGCGGCCGAG/g') and afterwards mapped back to the GATC motifs using the R DamID pipeline (see above; additional parameters: fasta.format="fasta"). Gaps between GATC motifs where calculated using the Biostrings R package (version: 2.36.4). The median distances between GATC motifs were plotted averaged by 25 kb bins. To calculate the distances for mappable GATC motifs, only GATC motifs mapped by the *in silico* reads were considered.

Figure legends

Figure 1

Experimental steps for DamID-seq analysis of protein-DNA interactions. **(a)** The DNA adenine methyltransferase (Dam) nuclear protein (protein of interest) fusion is expressed

in vivo at low levels. GATC motifs are methylated by Dam when they are located in close proximity to the protein of interest. A strain expressing *GFP::Dam* fusion (free Dam) is used as a control for DNA accessibility. **(b)** Either purified genomic DNA or entire lysed animals are used as starting material for DamID. Methylated GATC motifs are digested with *Dpn*I. **(c)** A double stranded adapter is ligated to the cleaved sites and fragments flanked by two G^{m}ATC motifs are amplified by PCR. After the PCR a smear between 500 bp and 2000 bp is visible. The first line on the gel after the ladder (Thermo Scientific™ GeneRuler 1 kb Plus DNA Ladder) shows buffer only as an empty control. The second line shows the negative control (20 N2 worms, no Dam fusion protein expressed) followed by the two separated lines with the characteristic smear from 20 hermaphrodite *Dam::lamin* worms. **(d)** Illumina sequencing Y adapters are ligated before HTS. The gel shows the resulting smear after limited library amplification by PCR for two 20 hermaphrodite *Dam::lamin* worms samples and one 20 hermaphrodite *GFP::Da*m worms control using the same ladder as in c). **(e)** Sequenced reads are processed for quality control before mapping to the genome and counting of methylated GATC motifs.

Figure 2

Required sequencing depth and achievable resolution **(a)** Percentage of mappable GATC sites found methylated versus number of valid DamID reads after QC for *GFP::Dam* (green) and *Dam::lamin* (blue) samples. The red line indicates one million DamID reads. **(b)** Effect of binning size on Pearson correlation between two *Dam::lamin* sequencing libraries (50'000 DamID reads each, low read number taken on purpose). While libraries are very well correlated at high bin size (>10 kb), the correlation is rapidly lost when bin size becomes smaller. **(c)** *Dam::lamin* Pearson correlation (at the single GATC level) between the full sequencing library and fractions of the same library for empirical determination of sufficient sequencing depth. Correlation is made for 100 kb genome segments. Black: 4 million reads library; blue: 1 million reads library; green 5% of the reads green. The error bars show the maximal and minimal value of the replicas and the

red line marks 0.95 correlation. **(d)** DamID-seq profiles with 100 kb bins made from 1 or 20 adult hermaphrodites animals, expressing either *Dam::lamin* or *GFP::Dam* fusions. **(e)** *Dam::lamin* profile with 100 kb bins normalized with a free GFP-Dam fusion. The enrichment for lamin interactions for the outer thirds of autosome arms are clearly visible (Ikegami *et al.*, 2010). Vertical red lines separate the chromosomes. The red line corresponds to previously published profiles using DNA as a starting material and microarray hybridization (Towbin *et al.*, 2012).

Figure 3

Adaptation of the methyl-tracer (m6A recognizing domain fused to GFP) system (Kind *et al.*, 2013) in *C. elegans* to visualize methylated regions in single cells.

(a) Schematic representation of the m6A-tracer system. **(b)** A wide-field partial projection image of an entire L3 stage worm co-expressing the lamin-dam fusion and m6A-tracer in a *glo-1* mutant background. **(c)** (top) Partial projection of nuclei in the pharynx of control L4 stage worms co-expressing the m6A-tracer. Single nuclei are represented in the bottom as either a z-projection (left) or individual slices (middle, right). **(d)** Partial projection of nuclei in the pharynx of lamin-dam L4 stage worms co-expressing the m6A-tracer, highlighting cell-type specific organization of lamina-associated chromatin. While in most nuclei the m6A-tracer shows a ring-like distribution (open triangles), some nuclei clearly display clustering of lamina-interacting chromatin (closed triangles). Single nuclei are shown at the bottom as either a z-projection (left) or individual slices (middle, right). Images **(c)** and **(d)** were acquired on a Spinning disk confocal microscope.

Figure S1

Theoretical considerations for DamID in *C. elegans* **(a)** Median distance in bp between GATC sites (black) and mappable GATC sites (blue) over all chromosomes averaged over

25 kb. The red vertical lines separate the chromosomes. The median over all chromosomes is indicated with a horizontal dotted line and the mean with a plain line **(b)** Boxplot of gap distances between all GATC sites, mappable GATCs using reads ranging from 25 to 87 bp as well as for a real sequencing library made from 20 animals expressing *Dam::lamin*.

Figure S2

Fragment length density for GATC sites mapped and not mapped in a *GFP::Dam* library, considering only mappable GATC. The fragment length distribution follows a Gaussian bell shape with its maximum around 500 bp for mapped reads reflecting the size limitation of the PCR and the purification steps excluding small fragments. In contrast the length distribution of fragments not mapped in all samples is bimodal with a high density below 100 bp. A further local maximum (around 5000 bp) represent the fragments which are too long to be amplified by PCR and/or sequenced using HTS.

Figure S3

Correlation plots of GATC read numbers for (a) one worm *Dam::lamin* using different PCR cycles after adapter ligation from 28 to 34 cycles, (b) 20 worms *Dam::lamin* replicas and (c) the correlation between samples (*Dam::lamin*) and control (*GFP::Dam*) using the same PCR cycle number. The correlation between 28 against 34 PCR cycles shows the same linear behaviour as the correlation with 28 against 30 cycles. As we do not observe a change in the behaviour of the correlation with increasing PCR cycle differences we can assume that PCR jackpot events are rare compared to *bona fide* amplification reflecting methylation frequency.

Figure S4

Intensity measurements of methyl tracer signal in *dam::lamin* or *gfp::dam* strains. (a) Intensity profiles were plotted for 5 individual nuclei across the orange dotted line in dam

fused to nuclear lamina as shown in the top image. Notice the intensity peaks at the nuclear periphery with a modest drop in the middle of the nucleus. (b) Same as (a) but for free diffusible dam. Contrary to the *dam::lamin* profiles, the signal peaks in the middle of the nucleus with multiple peaks representing patchy appearance. Profiles were generated using the 'Plot profile' function in ImageJ. All images were subjected to same thresholding.

Acknowledgements

The authors wish to thank members of the Meister laboratory for numerous discussions; Julie Campos for expert technical help, Mylène Docquier (iGE3 genomics platform of the University of Geneva) for NGS and advice. We kindly thank Dr. Jop Kind and Prof. Bas Van Steensel (NKI, Amsterdam) for sharing plasmids and technical expertise. *dam::lmn-1* and *gfp::dam* strains were a kind gift from Dr. Peter Askjaer (CABD, Seville). Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). The Meister laboratory is supported by the Swiss National Science Foundation (SNF assistant professor grant PP00P3_133744/159320), the Swiss Foundation for Muscle Diseases Research and the University of Bern.

Askjaer P, Ercan S, Meister P. 2014. Modern techniques for the analysis of chromatin and nuclear organization in C. elegans. WormBook : the online review of C. elegans biology: 1-35.

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, van Steensel B. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143: 212-224.

Gomez-Saldivar G, Meister P, Askjaer P. 2016. DamID of Nuclear Envelope proteins in C. elegans. In: Shackleton S, Collas P, Schirmer EC, editors. The Nuclear Envelope, Methods and Protocols: Springer.

Gonzalez-Aguilera C, Ikegami K, Ayuso C, de Luis A, Iniguez M, Cabello J, Lieb JD, Askjaer P. 2014. Genome-wide analysis links emerin to neuromuscular junction activity in Caenorhabditis elegans. Genome biology 15: R21.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453: 948-951.

Ikegami K, Egelhofer TA, Strome S, Lieb JD. 2010. Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. Genome biology 11: R120.

Kind J, Pagie L, de Vries SS, Nahidiazar L, Dey SS, Bienko M, Zhan Y, Lajoie B, de Graaf CA, Amendola M, Fudenberg G, Imakaev M, Mirny LA, Jalink K, Dekker J, van

Oudenaarden A, van Steensel B. 2015. Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. Cell 163: 134-147.

Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore WA, van Steensel B. 2013. Single-cell dynamics of genome-nuclear lamina interactions. Cell 153: 178-192.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10: R25.

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

Meister P, Towbin BD, Pike BL, Ponti A, Gasser SM. 2010. The spatial dynamics of tissue-specific promoters during C. elegans development. Genes Dev 24: 766-782.

Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B. 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nat Genet 38: 1005-1014.

RCoreTeam. 2013. R: A Language and Environment for Statistical Computing.

Sharma R, Jost D, Kind J, Gomez-Saldivar G, van Steensel B, Askjaer P, Vaillant C, Meister P. 2014. Differential spatial and structural organization of the X chromosome underlies dosage compensation in C. elegans. Genes & development 28: 2591-2596.

Southall TD, Gold KS, Egger B, Davidson CM, Caygill EE, Marshall OJ, Brand AH. 2013. Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. Developmental cell 26: 101-112.

Towbin BD, Gonzalez-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM. 2012. Step-Wise Methylation of Histone H3K9 Positions Heterochromatin at the Nuclear Periphery. Cell 150: 934-947.

van Steensel B, Delrow J, Henikoff S. 2001. Chromatin profiling using targeted DNA adenine methyltransferase. Nat Genet 27: 304-308.

van Steensel B, Henikoff S. 2000. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol 18: 424-428.

Vogel MJ, Peric-Hupkes D, van Steensel B. 2007. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nat Protoc 2: 1467-1478.

Figure 1.

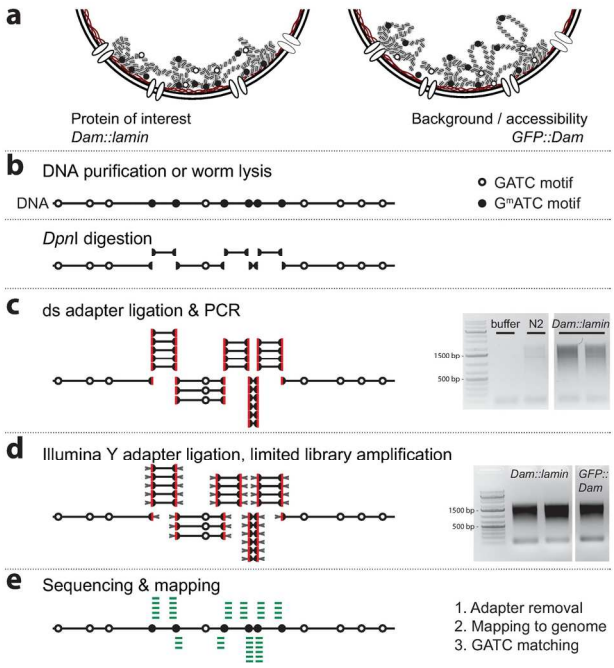

Figure 1
168x154mm (300 x 300 DPI)

Figure 2.

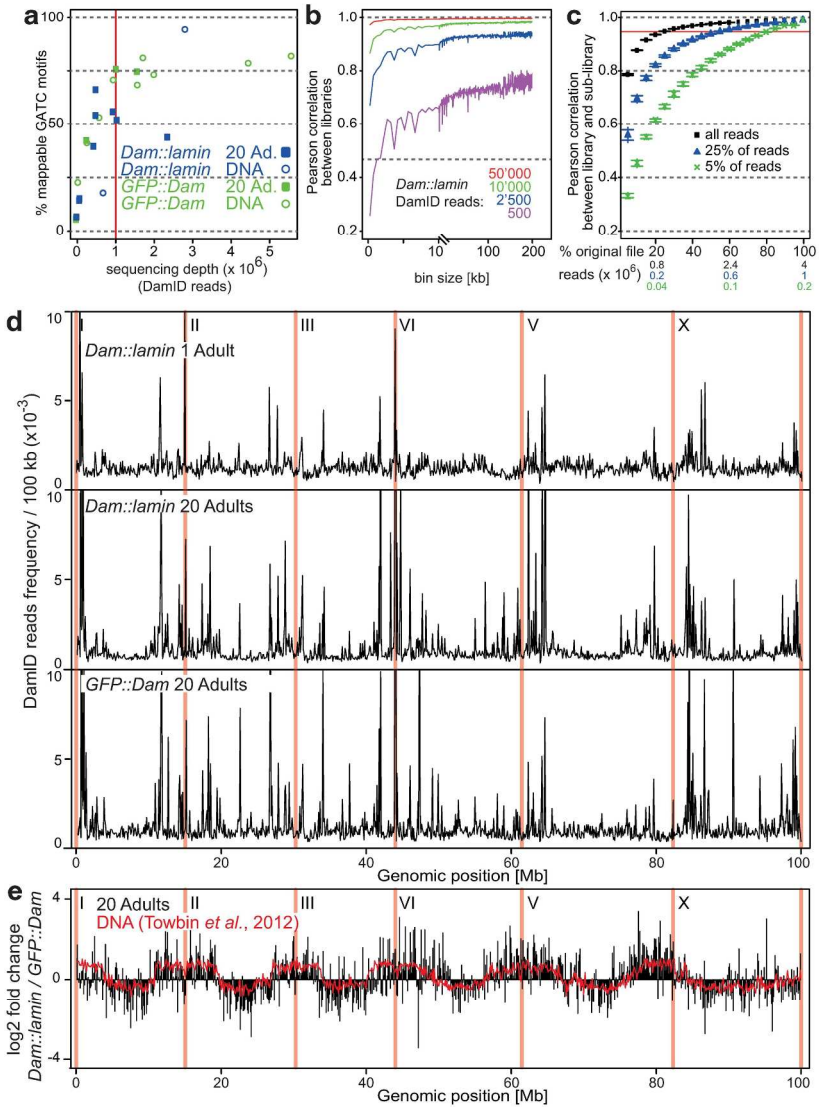

Figure 2
236x291mm (300 x 300 DPI)

Figure 3.

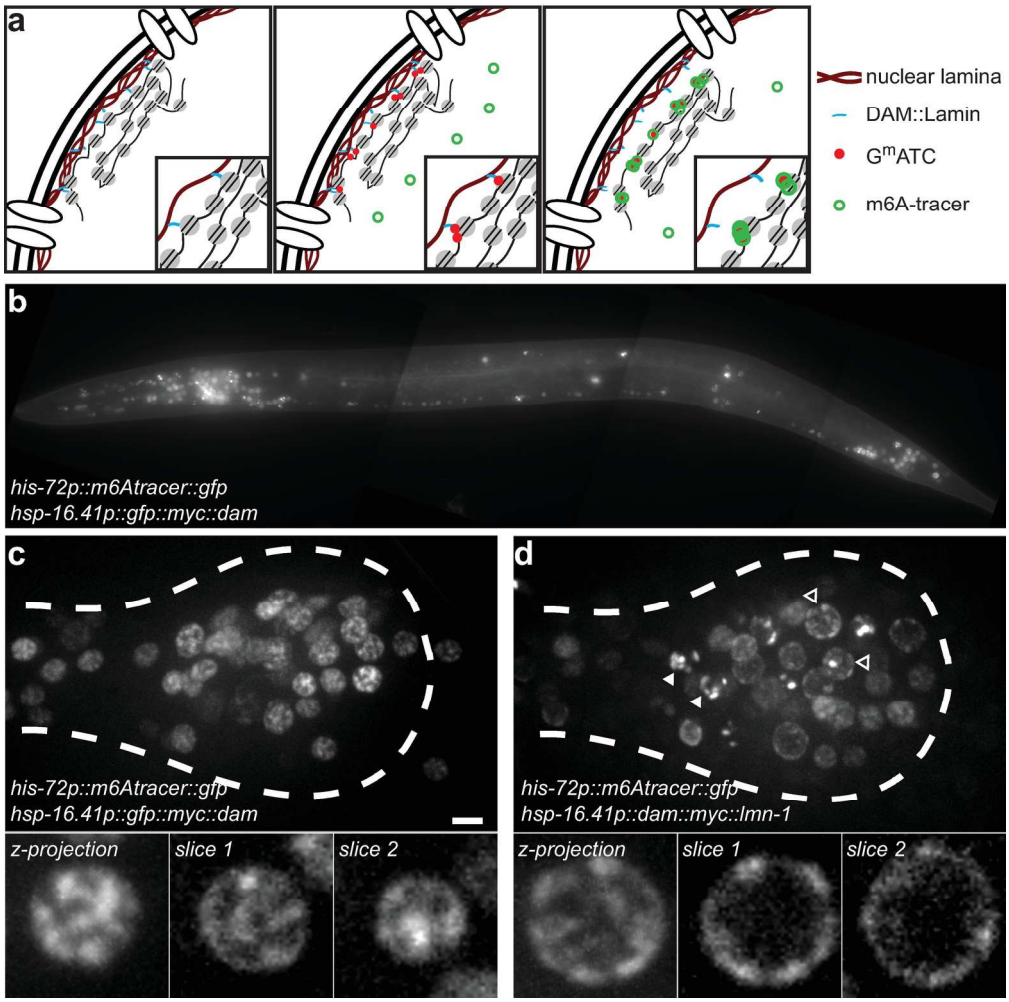

Figure 3
183x199mm (300 x 300 DPI)

Figure S1.

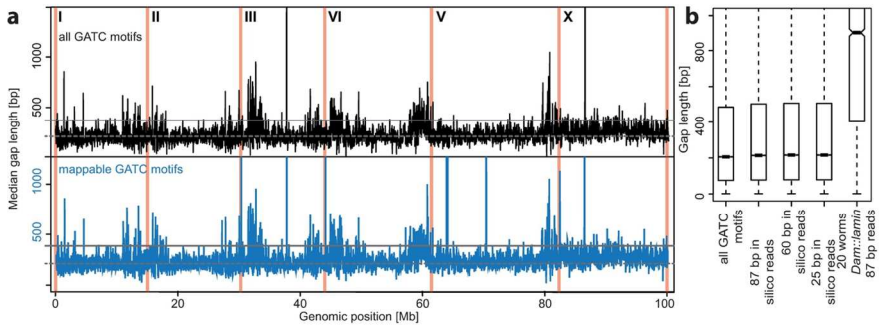

Figure S1
132x75mm (300 x 300 DPI)

Figure S2.

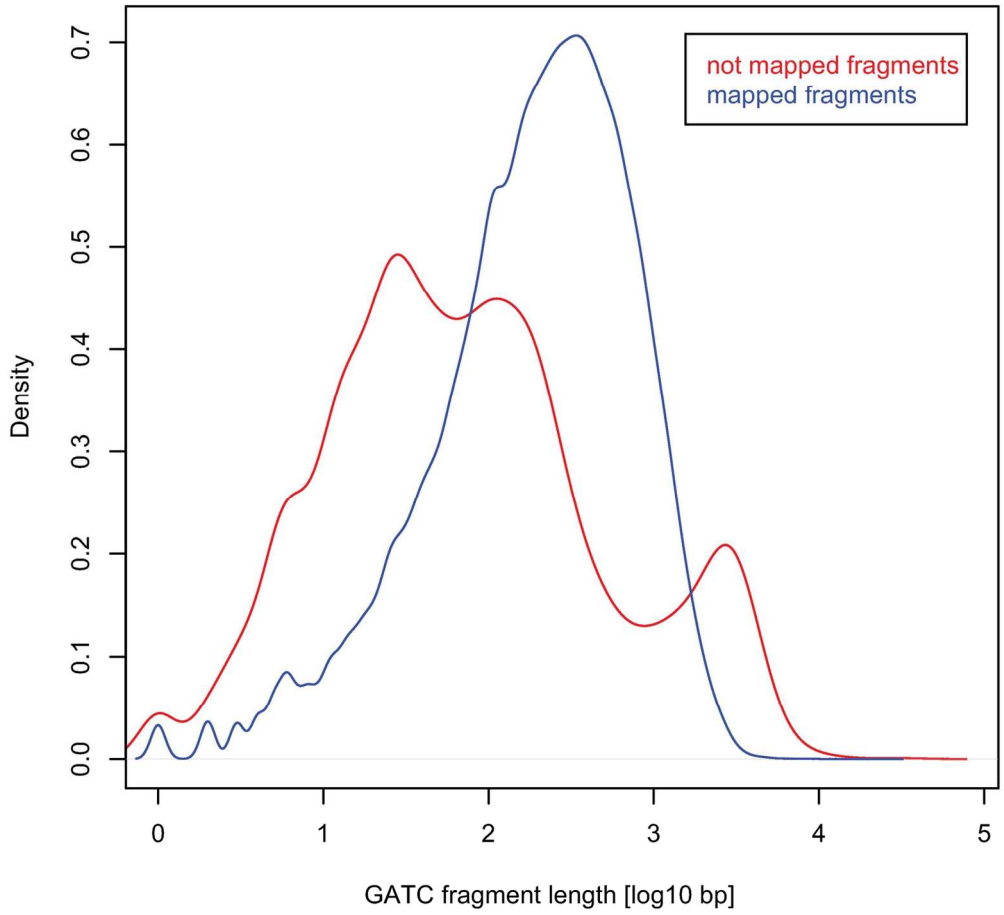

Figure S2
135x139mm (300 x 300 DPI)

Figure S3.

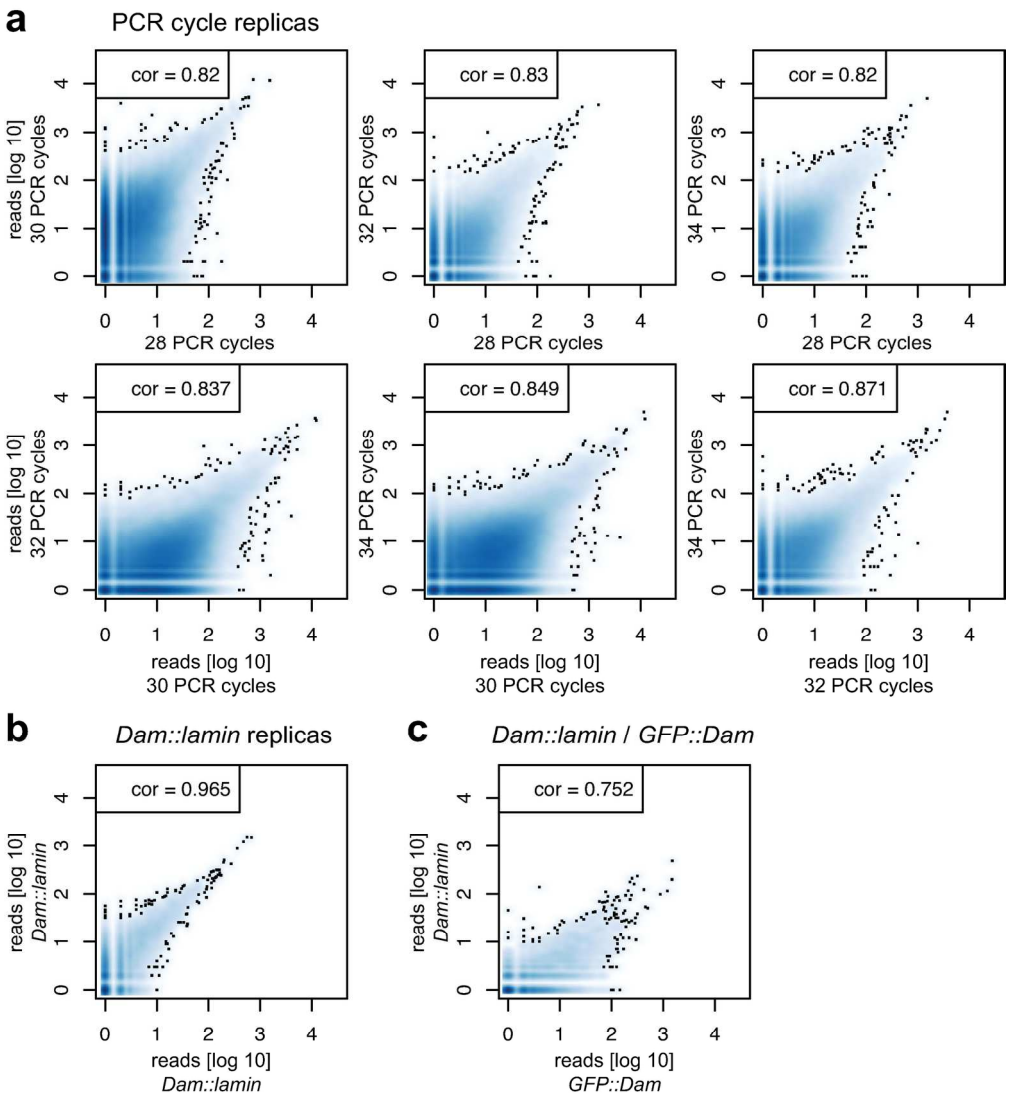

Figure S3
169x201mm (300 x 300 DPI)

Figure S4.

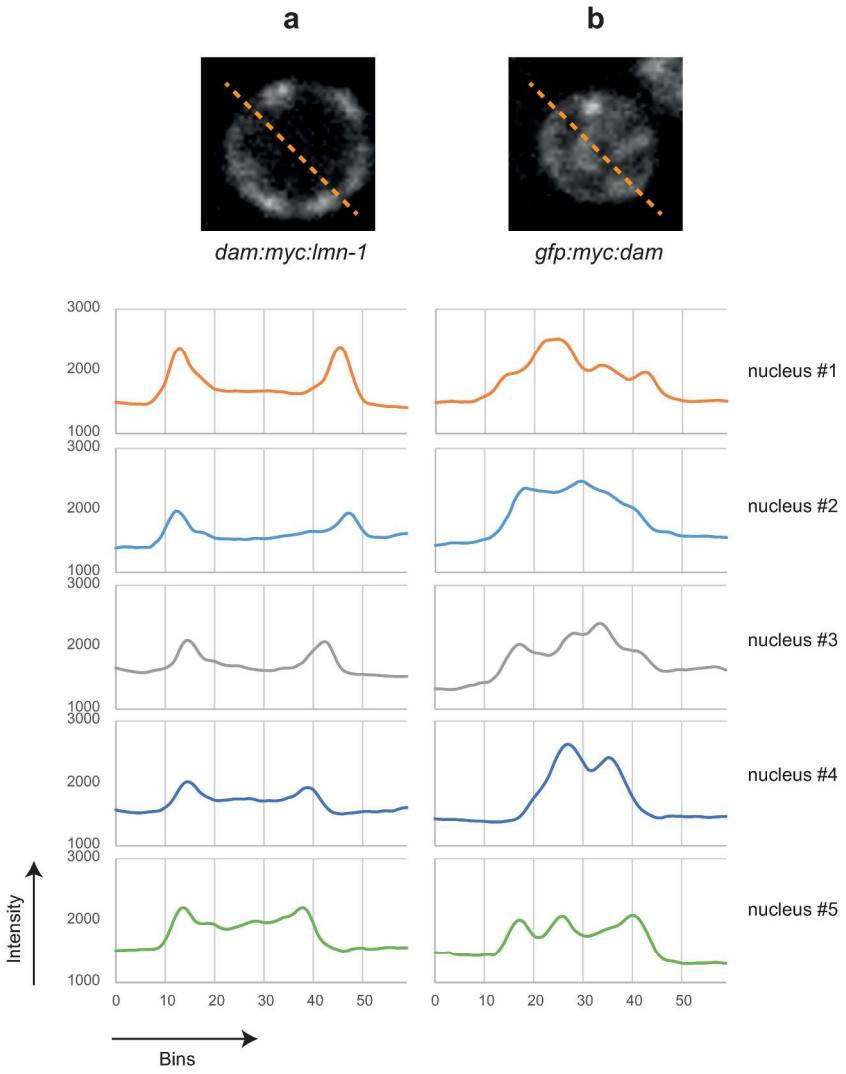

Figure S4
226x324mm (300 x 300 DPI)